REPORT RESUMES

ED 010 525 24

A COMPUTER ANALYSIS OF FICTIONAL PROSE STYLE.
BY- KROEBER, KARL
WISCONSIN UNIV., MADISON
REPORT NUMBER CRP-3067 PUB DATE OCT 66
REPORT NUMBER BR-5-0234
CONTRACT OEC-6-10-015
EDRS PRICE MF-$0.18 HC-$5.04 126P.

DESCRIPTORS- *DATA PROCESSING, ELECTROMECHANICAL AIDS, *ENGLISH LITERATURE, LITERATURE PROGRAMS, *SYNTAX, LANGUAGE PATTERNS, *VOCABULARY, WRITING SKILLS, LINGUISTIC PATTERNS, *FICTION, CHARLOTTE BRONTE, EMILY BRONTE, ANNE BRONTE, GEORGE ELIOT, JANE AUSTEN, MADISON, WISCONSIN

FUNDAMENTAL CHARACTERISTICS OF FICTIONAL PROSE STYLE WERE STUDIED THROUGH SYSTEMATIC AND OBJECTIVE ANALYSES OF NOVELISTIC SYNTAX AND VOCABULARY. SAMPLE PASSAGES FROM THE MAJOR NOVELS OF JANE AUSTEN, THE BRONTE SISTERS, AND GEORGE ELIOT AS WELL AS NOVELS BY 13 OTHER AUTHORS WERE ANALYZED. INFORMATION ON PASSAGE SENTENCES, CLAUSES, AND WORDS WAS CODED AND TRANSFERED TO MAGNETIC TAPE. STATISTICAL TESTS WERE RUN ON THE DATA, AND FREQUENCIES OF SYNTACTIC PATTERNS AND VOCABULARY PREFERENCES WERE PRINTED OUT. THE PRIMARY CONCLUSIONS OF THE STUDY WERE--(1) IT IS NOT POSSIBLE TO DEFINE THE STYLE OF ANY NOVELIST THROUGH SIMPLE STATISTICAL ANALYSIS OF HIS GRAMMAR OR HIS WORD CHOICE, (2) NOVELISTIC STYLE CAN BE SATISFACTORILY IDENTIFIED ONLY IN TERMS OF MULTIPLE FACTORS, MANY OF WHICH GO BEYOND THE LEVEL OF SYNTAX AND VOCABULARY, AND (3) FURTHER SYSTEMATIC STUDY OF FICTIONAL PROSE STYLE SHOULD BE BASED ON AUTOMATED ANALYSIS OF TEXTS, AS THE HUMAN ANALYSIS OF TEXTS REQUIRES AN EXORBITANT AMOUNT OF TIME. (AL)

ED010525

CRP 3067
(5-0234)

FINAL REPORT

Project No. 3067

Grant No. OE-6-10-015

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A COMPUTER ANALYSIS OF FICTIONAL PROSE STYLE

October 1966

U.S. DEPARTMENT OF

HELATH, EDUCATION, AND WELFARE

Office of Education

Cooperative Research

A COMPUTER ANALYSIS OF FICTIONAL PROSE STYLE

Project No. 3067
Contract No. OE-6-10-015

Karl Kroeber
October 1, 1966

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Wisconsin
Madison, Wisconsin

TABLE OF CONTENTS

Introduction.. 1
Method.. 5
Results... 21
   Part 1 (Contrast).................................. 22
   Part 2 (Totals).................................... 32
Discussion.. 69
   Contrast... 71
   Totals... 85
Conclusions... 93
Summary... 96
Bibliography..101
Appendix A, List of Novels............................A-1
Appendix B, Macro-marking Procedures..................B-1
Eric Resume Form......................................C-1

INTRODUCTION

This project focused on the problem of defining "style" in the novel. The significance of the problem is dramatized by the absence of any accepted definition of the term "style" as applied to fictional prose -- though many have been offered over the years. But this is only a specific manifestation of a broader difficulty in humanistic scholarship: no one has succeeded in establishing as uncontroversial and clear-cut meaning for the word "style" in any of the arts, although humanists agree that "style" is essential to all art and art history. (Good recent summaries of the various kinds of definitions proposed for style in literary studies are to be found in the works by Lodge, Milic, Sayce, and Ullman listed in the bibliography.)

A basic assumption upon which this project was founded was that a good method of working toward some generally acceptable definition of "style" would be to single out one area of the arts, literature, then delimit a relatively small and relatively coherent formal part of that area, the English novel, and, finally, to select one epoch in the history of that part, the nineteenth century, for intensive study. My thought was that if I could describe with some objectivity and in some detail characteristics of the "style" of the nineteenth-century British novel I would lay the groundwork for the development of systematic studies into the nature of style in other areas of literature and the other arts.

My method, however, depended upon several presuppostions. One of these was that the problem of style in literature is distinct from, though of course related to, the problem of "authorship." The question, for example, of who wrote under the name of "Junius" is not primarily a stylistic question but primarily a question of authorship (see Ellegard, Bibliography). My study is aimed at determining stylistic characteristics of novels rather than novelists, if such a distinction can be sustained. "Style" so conceived is not alone what distinguishes one literary work from all other literary works but also what simultaneously associates it relatively closely with some works, less closely with others, and very tenuously with still others.

This presuppostion explains why the study of style in literature is so difficult. The student must describe characteristics which are simultaneously unique to a given work and shared by it in varying degrees with other works. Temporal relation is an obvious form of "sharing," recognized implicitly at least, whenever we speak of, say, Elizabethan Drama or Augustan Poetry. Genre is another form of association, implied by terms such as "tragedy" and "the novel."

An obvious corollary to this assumption is that a stylistic study, as distinct from an investigation into authorship, must be concerned with the work of more than a single artist. Put into its most extreme form, this means that there cannot be a meaningful study of the literary "style" of a single writer in isolation from other writers. A study of the style of Milton, for instance, conducted without reference to any writers associated with him in time, through genre, by philosophy or religious conviction, and so forth, would be a contradiction in terms. In fact, there are no such studies. The very process of describing the idiosyncratic can not be carried out without reference to that from which the idiosyncratic is distinguishable. But there have been very few studies which have directly attacked the problem of literary style as one in which multi-related characteristics must be defined as simultaneously discriminative and associative.

One reason there have been few such studies is that the labor involved is enormous and the task an intricate one. Merely to describe with precision, to use the previous example, some of the characteristics by which readers recognize Milton's poetry as distinctively "Miltonic" requires both extensive knowledge and subtle insight. Merely to describe some of the fundamental conventions and purposes of, say, "Elizabethan Drama," demands the combination of extensive learning and developed critical acumen. To unite these endeavors, to join the precise and detailed understanding of one writer's unique use of language with a broad comprehension of how that uniqueness relates his work to the uniquenesses of works written by others, is both a large and a complicated enterprise. Yet it has to be undertaken if we wish to discuss and evaluate "style" meaningfully.

I assume, in other words, that what is needed is a method for systematic studies of literature which are simultaneously intensive and extensive, which are critically two-dimensional, at once singling out unique characteristics and establishing relationships among them within a context of definably connected works.

It occurred to me that a computer might contribute to the development of such a method, because a computer is useful for organizing large masses of detailed data, can compute rapidly, and can be programmed to figure quantitative relationships neatly and swiftly. Style certainly can not be reduced to quantifiable elements only, but the qualitative aspects of style are accompanied by features which do lend themselves to systematic measurement. The judgment that a characteristic of a given work of literature is unique is a judgment which, theoretically at least, is measurable. The example is used because presumably one virtue of computer analyses of literary data is that the machine can more easily than a human being isolate uniquenesses of detail. It is in its power to locate the unique amidst the obscuring mass of detail in any sizable literary work that the computer may serve the literary scholar as the microscope has served the scientist.

I assume, however, that a systematized study of style which utilizes mechanical means of arranging and analyzing literary data should not become an investigation of language rather than literature. Since literary art is impossible without language, there is a temptation, virtually irresistible to linguists, to regard explanations of systems of language as the key to defining the systems of literary style. The literary artist uses language, and the student of his style should be interested primarily in his manner of use, not that which he uses. Thus the form of the English novel changed far more radically between 1741 and 1941 than did the English language, and to the student of novelistic style these changes in form must be more significant than the relative stability of the language.

This point deserves stress because it provides a basis for distinguishing the work of this project from most other recent stylistic studies which treat of details of language in literary texts. Moreover, intrinsic to the problem to which I have addressed myself is our lack of information about it. We simply do not possess a substantial corpus of analyzed novelistic prose upon which to found judgments, from which to erect hypotheses, against which to test and qualify our intuitive critical perceptions, and with which to develop more precise scholarly terminology. In a very real sense the most fundamental obstacle to the study of style in sophisticated novelistic prose is the absence of any coherently organized body of data which enables scholars and teachers to compare novels concretely and in discriminating detail. The original aim of this project was to supply a substantial body of such information as the foundation for more advanced work later.

The specific objectives of this study were summarily listed in the original proposal as follows:

1. To describe and to define the syntactic characteristics of the fictional prose style of five important nineteenth-century novelists (Jane Austen, Charlotte, Emily, and Anne Brontë, and George Eliot).

2. To describe and define the characteristic vocabulary found in the fictional prose of these five novelists.

3. To define changes and developments within each of the five authors' styles through study of the vocabulary and syntax of all the novels of each author.

4. To define special relationships which may exist between the styles of the three sisters Charlotte, Emily, and Anne Brontë.

5. To define relations between the style of Jane Austen, her novels having been published originally between 1811 and 1818, the Brontës, their novels having been published originally between 1847 and 1857, and George Eliot, whose novels originally were published between 1858 and 1876.

6. On the basis of the above to begin to define some factors characteristic of prose style in the nineteenth-century British novel.

7. To attempt to establish the relationship of dictional and syntactic characteristics of novelistic prose style to "macro-syntactic"

characteristics, that is, features of plotting, characterization, and the like.

Beyond these specific aims, however, were the fundamental, underlying objectives of the project:

1. To compile an organized body of information, of a kind never before collected, on the grammar and vocabulary of nineteenth-century novelists which might serve as a basis for further, more detailed, and more illuminating investigations of novelistic prose and for experiments in new techniques of teaching both literature and composition.

2. To lay a groundwork for systematic, relatively objective, and cumulatively-rewarding analyses of the nature and value of fiction.

The presupposition underlying these objectives was that both the style of a particular novelist and the style of a literary epoch are temporal processes that can be defined satisfactorily only as patterns of shifting multiple relationships. Unlike those who have sought to solve problems of authorship, I hoped not to isolate a few separately distinguishing characteristics but to bring into focus some underlying systems of prose manipulation which simultaneously associate and distinguish novelists practising sequentially during a specific literary epoch.

Since no work of this kind had been attempted previously, it seemed wise to collect and order as much data as possible in an easily comprehensible form without attempting anything statistically sophisticated. This decision probably precluded any sensational results. But, because I unquestionable underestimated the bulk and the complexity of the data that was actually collected, I do not believe that I was wrong to aim for a simple methodology.

METHOD

Selection of the principal authors to be studied (Jane Austen, the Brontës, George Eliot) was based on the following criteria.

a: My special competence is nineteenth-century British literature. I have studied and taught the works of these novelists for several years.

b: The total number of novels written by each of the authors is small enough to make feasible the comprehensiveness desired.

c: Two of the novelists at least, Austen and Eliot, are recognized as of the first rank.

d: The Brontës are customarily regarded as holding a rather special position in the history of the novel and therefore provide means of testing whether the "central tradition" is as clearly marked as is often asserted.

e: The Brontës provide a virtually unique opportunity for the study of - literally - sister authors.

f: The authors span a major period in the history of the novel. One must move backward more than thirty years before the publication of _Pride and Prejudice_ (1811) to find a novel of comparable excellence, and Eliot's work is frequently regarded as the climax of the Victorian novel. The authors, nevertheless, belong to distinctive sub-periods: if Eliot is "high Victorian," the Brontës are "early Victorian," and Jane Austen's work falls in the Romantic period. Moreover, the Brontës knew Austen's work, but began writing after Austen's career was finished, and the same is true of Eliot's relation to the Brontës. In other words, these novelists form a clear temporal sequence.

Selection of other novels to be analyzed was based on my desire to have at least minimal representation of the entire span of British fiction from the 18th to the 20th centuries as a context for the main novelists studied. Novelists were selected, moreover, on the basis of particular relationships to the central foci of study. Thus, for example, five novels by Dickens were analyzed because of his extreme importance to the history of the nineteenth-century novel, Burney's and Woolf's novels were included because the authors were women, and so forth.

Because of the objectives of the project it was necessary to utilize relatively small samples from a good many novels rather than relatively large samples from a few novels. Samples analyzed (description of actual samples will be found in appendix A) were of three types.

1: "Block" sample: all sentences from about 10-20 continuous pages (including at least one chapter) near the mathematical center of the novel.

(In nineteenth-century novels this is usually an important stage in plot development and most often involves several major characters.)

2: "Random" sample: units of five consecutive sentences on pages chosen by means of a table of random numbers from those portions of the novel not covered by a "block" sample.

3: "Special" sample: a sample the size and nature of which varies because it is selected to test a particular intuitive judgment or a specific hypothesis arising from study of the other kinds of samples, or to represent a special kind of prose, narrative, dialog, etc.

It should be noted that more than one sample was analyzed for several novels. Where this occurred cumulative statistics could be developed. But creation of these is for some items difficult, because the process of analysis was changed after the project was launched. The procedures described below were used predominantly. But samples from seven novels were analyzed according to a slightly different system, so accumulation of results was not always possible.

Before describing in detail the procedures of analysis, I should like to make explicit two fundamental principles which determined the development of these procedures. First, complete objectivity in analysis was never my aim. The rigorous impersonality of, say, a mathematical demonstration would be inappropriate in a literary study. I was quite willing, therefore, to introduce into the process of my analyses elements such as a distinction between "concrete" and "abstract" nouns and between verbs of "physical" and "psychic" action, where personal decisions by the analysts would necessarily enter. Also I used even more analysts than were necessary so as to obtain a vide variety of subjective decisions. In the long run, I believe, systematic studies of literature will be valuable insofar as they do not try to escape from personal responses but incorporate appropriate subjectivity into their systems.

The basis of my grammatical analyses was traditional, "analytical" grammar, most of the distinctions in the system being based on the well-known handbook of Porter Perrin. I used traditional grammar for several reasons. It is closest to the grammar known by the authors I was studying. It is familiar to all of the analysts. The systems developed by modern linguists are not so well known, and my study of them led me to believe that, however promising the future of these new techniques, at present they are not sufficiently developed to provide a realiable basis for work such as mine. Much of my material could be readily converted to the terminology of, say, structural linguistics, and the computer records for each word, clause, and sentence are arranged so as to permit both the addition of new information and a re-classification of the data. In any event, the system used is not so important as the fact that it is comprehensible, and my arrangement will be readily understandable to anyone who had had a high-school course in either grammar or composition.

OPERATING PROCEDURES

Sample passages in the novels to be studied are marked off. The text of the sample is then typed on prepared sheets in columnar form, one word or mark of punctuation per line. Either I or one of my assistants then goes through these sheets indicating grammatical or semantic information about each word, clause, and sentence in a numerical code (see below) that a computer can deal with rapidly and efficiently. This "code" consists essentially of assignment of a column and number to each piece of information, such as sentence type, part of speech, etc. The completed information sheets are then given to a keypunch operator who punches the information on cards, one card for each word, mark of punctuation, clause, and sentence. Card images are transferred to magnetic tapes and from these tapes data files are constructed on another magnetic tape. Most of our work has been done on a Control Data Corporation 3600, which processes the card-images sequentially, prints these out, thus providing a permanent record and a means for error correction, then performs various analyses of the data and prints out the results.

Marking procedures.

STEP 1 - In the novel itself

In pencil in the left-hand margin, units (paragraphs) are numbered consecutively on each page. If unit #5, for example, continues from the bottom of page 113 to the top of page 114, "unit 5, page 113" is regarded as ending at the conclusion of the "transitional" sentence, part of which is on page 113, part on page 114 (or, if this transitional sentence is very long and extends far down on page 114, at the first semicolon or colon). In a case like this, the first unit on page 114, beginning after the "transitional" sentence, will be marked "0," and unit 1 will begin at the first indented paragraph. Herein lies the reason why we use the word "unit" instead of "paragraph." A list of all the characters--speaking as well as spoken of is made up, each character, including the author being assigned, a two-digit number--01, 02,...09, 10, 11, 12, and so on. In some books the author may appear as both narrator and speaking character--e.g., David Copperfield. In such cases he is given two different numbers.

STEP 2 - Proofreading the white sheets

The specially printed white sheets are consecutively numbered by means of the encircled numbers at the top. The number next to the word NOVEL indicates the number of the novel in this project.

The text of the novel on the white sheets is proofed and checked against the text in the novel itself. Special attention is paid to the accuracy of the punctuation--each mark of punctuation has its own line. An asterick appears behind each personal proper name; place names and adjectival forms are not so marked. Two or three spaces are left between clauses.

STEP 3 - PAGE and UNIT designations

The printed "PAGE" and "UNIT" at the top of the white sheets signal that page and/ or unit numbers of the novel begin at that point.

Page numbers are recorded in columns 6, 7, and 8, unit numbers in columns 5 and 6.

Page 54, Unit 1

column	5	6	7	8
PAGE			5	4
UNIT		1		

Page 378, Unit 1

column	5	6	7	8
PAGE		3	7	8
UNIT		1		

Page 378, Unit 12

column	5	6	7	8
~~PAGE~~				
UNIT	1	2		

Note that in the third example PAGE is crossed out. The words PAGE and/ or UNIT are to be crossed out unless a new page or unit of the novel in fact begins at the top of a white sheet.

STEP 4 - Sentence and clause designations

The printed SENT and CLAU at the top of the white sheets signal that sentences and/or clauses begin at that point. As with PAGE and UNIT, unless the SENT and CLAU actually signal the beginning of a sentence or a clause at that point, a line is drawn through one or both.

A clause might very well begin elsewhere on the white sheet. If so, CLAU is written in columns 1, 2, 2 and 4 directly above the first word of each such clause. If by chance a sentence begins elsewhere on the sheet, in columns 1, 2, 3 and 4 SENT is written in above CLAU.

At this stage in the marking, nothing further is done with SENT's. However, in each sentence clauses are numbered consecutively, using column 6 (and 5 if necessary) behind CLAU. We mark from the right--column 6 for one-digit numbers, columns 5 and 6 for two-digit numbers. Sentences are not so numbered.

Sometimes a clause may be interrupted by one or several separate clauses. In such a case, all of the "pieces" of the original clause must be marked CLAU and all given the same number in column 6 (and 5). The interrupting clauses continue to be numbered consecutively.

For example, suppose the first clause of a sentence is broken in two by two interrupting clauses, and is followed by yet another clause. The first clause up to the first interruption is marked 1 (in column 6), the first interrupting clause 2, the second interrupting clause 3, the other "half" or "piece" of the original clause after the interruptions again 1, and the last clause 4.

STEP 5 - PUNCTUATION designations

On the far left side of the white sheets, all marks of punctuation are identified using the following code. Note that the most common marks of punctuation, periods and commas, are not specially indicated.

+A	period in an abbreviation
+B	name of person or place indicated by initial letter and dash
--	dash as word hyphenator
+D	dash as sentence terminator
+I	dash as sentence initialization (e.g. quotation in French)
+-	dash as comma in sentence
+S	semicolon
+C	colon
+E	exclamation mark
+Q	terminal question mark
+R	initial inverted question mark (e.g. question in Spanish)
+(	left parenthesis
+)	right parenthesis
+1	
+2	1-3 dots used as pauses and not periods
+3	
+4	left double quotation mark
+5	left single quotation mark
+6	right single quotation mark
+7	right double quotation mark
+P	double punctuation
+N	ampersand

STEP 6 - SPEAKER designations

Columns 40ff. are coded to tell us such things as which character is speaking, soliloquizing, quoting, and so on, as well as whether he is entering or leaving. In these columns the appropriate character number is written down, prefixed by the following letters where applicable:

A	character begins to speak	
S	character ceases to speak	
M	character begins soliloquy	
N	character ends soliloquy	
Q	quotation begins (direct address--number with character who quotes)	
Z	quotation ends	
E	character enters	author gives specific indications, does not simply change scene.
L	character leaves	

Character identifiers and relevant prefixes may be placed opposite SENT, CLAU, or any appropriate word or mark of punctuation. A's and M's and the like are logically placed after SENT or CLAU and *before* any punctuation, while S's and N's logically belong opposite final punctuation marks. An author may be the only speaker for, say, two pages of a novel, and hence for, say, 39 white sheets. Thus, at the top of sheet 1, opposite SENT in columns 40, 41, and 42, is written A13. S13 will go after the period on page 39. Within these 39 pages, a character can enter or leave, and the E or L and character code number can be placed opposite any relevant word telling of such action.

The beginning of a quoted speech implies, of course, both the end (S) of the author's speech and the beginning (A) of the speech of the character. The interruption of a quoted speech by the author--"Yes I will," said John, "if you say so"--requires an A and an S for John, an A and an S for "author" (as 'speaker' of *said John*), another A for John, and so on. The interruption of a quoted speech by another character, or back to back quotations of different characters, of course require A's and S's for each character involved.

STEP 7 - Detailed marking of sentence (SENT)

In Column 7 (Construct type)

1 simple
2 compound
3 complex
4 compound-complex
5 fragment
6 fragment with clause (s) or clause with fragment (s)

Note: quoted matter introduced by an author's comment like "he said" is regarded as grammatically independent of the "he said." Hence a quotation, even a fragmentary one, is not to be regarded as a noun clause completing the verb "said." The quote is one separate clause, the "he said" another.

In Column 8 (Mood type)

1 declarative
2 interrogative
3 exclamatory
4 imperative
5 mixed

In Column 9 (Discourse type)

1 descriptive (obvious effort to describe, to paint, or set scene)
2 narrative (the most frequent class, the plot line)
3 direct discourse (dialog in quotation marks)
4 indirect discourse (he said that)
5 direct discourse together with author speech ("Yes I will," said John, "if you say so," as he left the room.)
6 other expository (e.g., soliloquy)
7
8
9 letter
0 expository (author leaves plot line to philosophize, comment, preach, or teach--e.g., introductory chapters in Tom Jones)

STEP 8 - Detailed Marking of Clauses (CLAU)

Clauses are either independent (main) or dependent (subordinate). If a clause is independent, a 1 is placed in column 7 and the work is done. (If a clause is fragmentary--usually a direct quote--a 3 is placed in column 7.) If a clause is subordinate, a 2 is placed in column 7, and in column 8 the appropriate number telling the kind or function of the subordinate clause is assigned.

Column 7

1	independent (main)
2	subordinate (dependent)
3	other (fragment)

Column 8 (Type of subordinate clause)

1	noun clause - subject of sentence
2	noun clause - other
3	adjective clause - obviously restrictive
4	adjective clause - other
5	adverb clause - time
6	adverb clause - place
7	adverb clause - cause
8	adverb clause - other
9	conditional clause

STEP 9 - Labelling parts of speech

Each word is identified as a certain part of speech (noun, pronoun, verb, etc.) according to the following code:

	Column 5	Column 6	Column 7
1	noun		
2	pronoun (and in col. 8 mark a 2 if subject, 3 if anything else)		verbal--infinitive (to kill see #6 below)
3	verb	adverb (even those closely connected with verbs)	verbal--present participle
4	main part of compound verb		verbal--past participle
5	auxiliary part compound verb (had been, have been going		verbal--gerund
6	verbal (requires an additional mark in col. 7--#2, 3, 4, or 5)	Preposition	"to" in infinitive constructions (and auxiliary members of the construction: for example, the "be" in to be killed)
7	main part compound verbal (John, having acted rashly, quit. An infinitive construction is not a compound verbal--see col. 7, #2.)	conjunction	interjection
8	auxiliary part compound verbal (having acted)	contraction of noun/pronoun and verb (place number behind both)	foreign words
9	adjective	contraction of verb and adverb (place number behind both; the 9 replaces the 3 in this column for adverb)	

STEP 10 - Parts of speech themselves--their sentence functions

Remaining marking work consists in description of the various parts of speech and their roles in the sentence.

A. NOUNS

Column 8 (subject or not)	Column 9 (number)	Column 10 (kind of noun)	Column 11
1	1 singular	1 concrete person	A 1 is marked if a common noun is used where a proper noun would be possible. Otherwise left blank. (E.g., a man in a conversation with an acquaintance refers to him as "friend" instead of "Steve.")
2 subject of sentence or clause	2 plural	2 concrete place	
3 other function	3 collective	3 concrete thing	
	4 undetermined	4 abstract quality or condition (personal traits, virtues, vices)	
		5 abstract actions (you can "see" movements: "went for a walk")	
		6 abstract complex, subjective, connotative judgments, ideals, emotions, feelings (love, hate, freedom, motherhood	
		7 proper names--persons and places	
		8 abstract collective names and concepts--most abstract words which do not fit any other category	
		9 time words--days, months, and any word with a time reference	

B. ADJECTIVES

Column 8 (degree)

1 positive
2 comparative
3 superlative
4 not applicable (i.e., the adj. is such that it can have no degree other than itself--proper names used as adj., "first")

Column 9 (nature of adj.)

1 descriptive adj. of measure
2 descriptive adj. of quality
3 limiting--definite article (the)
4 limiting--indefinite articles (a, an)
5 limiting--demonstrative (this, that)
6 limiting--numerical (second)
7 limiting--pronominal (e.g. his, hers)
8 other--definite (e.g. proper names used as adj.)
9 other--indefinite

C. ADVERBS

Column 8 (degree)

1 positive
2 comparative
3 superlative
4 not applicable

Column 9 (nature)

1 tells how or how much
2 tells when
3 tells where
4 tells why
5
6 negative (no, not)
7
8 expletive (yes, indeed, certainly)
9 intensifier (very, too--must be followed by adj. or adv.)
0 introductory filler word, or explicative: "there is" and "it is"

Column 10 (function)

1 simple (modifies a verb, adj. or adv.)
2 clausal (modifies whole clause or sentence)

D. CONJUNCTIONS

Column 8 (nature)

1 coordinate conjunction linking clauses — and, or, nor, but, yet, for
2 coordinate conjunction not linking clauses
3 correlative
4 conjunctive adverb (however, therefore, etc.)
5 subordinating conj. (single--while, since, etc.)
6 subordinating conj. with another (in text or implied)
7 punctuational conj.

E. VERBAL

Column 10 (function as part of speech)

1 substantive
2 modifier
3 other

Column 11 (syntax function)

1 as subject of sentence
2 as subject of clause (otherwise leave blank)

F. VERBS

Column 8 (voice)

1 active
2 passive
3 copulative (verbs other than "to be" used copulatively are marked either 1 or 2 in col. 8, and 3 in col. 9--verbs such as seem, feel, became, appear, grow, look, continue)

Column 9 (use)

1 transitive
2 intransitive
3 copulative
(verbs introducing quoted passages--"he said"--are regarded in this study as intransitive)

Column 10 (tense)

1 present (requires additional mark in col. 13)
2 past
3 future
4 present perfect
5 past perfect
6 future perfect
7 progressive present
8 progressive past
9 progressive future

Column 11 (tense cont'd)

1 progressive present perfect
2 progressive past perfect
3 progressive future perfect
4
5
6
7
8 modal present
9 modal past

Column 12 (tense cont'd)

1 modal present perfect
2 modal past perfect
3
4

NOTE: For purposes of this study there are three classes of verbs--the "simple," the "progressive," and the "modal." The simple is no problem. The progressive verbs imply a continuing action in their respective tense--"he had been talking on and on"; in this category also belong the emphatic forms involving some form of DO. The modals are those which, in most foreign languages, are called subjunctives or conditionals: should, would, could, can, may, might, and must.

Column 13 (types of present)

1 universal (eternal time; true for all times--"women are fickle")
2 repetitive (no stated time, but an action is continuing to go on)
3 historical (action really occurred in past, but is expressed with present verb form--"I am walking past his pool hall, when all of a sudden...." (Runyon)
4 regular (actual present time)
5 imperative

Column 14 (verb action type)

1 physical action (i.e., bodily, involving movement)
2 psychic or mental action (e.g. thinking, wishing)
3 other
A 3 is marked for all verbs unless a 1 or 2 classification is obvious and clear.

The information as coded according to the description above was keypunched on to IBM cards, one sentence, clause, and word with associated information per card. The card data wastransferred to magnetic tape, first in the form of card images, then in a condensed form. A print-out of the condensed form giving sequence numbers for each record was provided for proof-reading and a permanent record. A computer program then performed a series of counting, sorting, grouping, and analyzing operations and printed out the results of these. The operations performed are illustrated in the RESULTS section of this report and are discussed in the following section.

Another computer program compiled vocabulary lists and listed all sentences in a sample in sequential order with a count of the number of words in each sentence, its sequence number in its paragraph, and the "construct," "mood," and "discourse" type assigned to it. The program compiled vocabulary lists only for words which had been categorized as follows (in the final print-out there was a separate list for each category indicated):

noun - concrete of person
noun - concrete of place
noun - concrete thing
noun - abstract quality
noun - abstract action
noun - abstract idea
noun - abstract concept
noun - abstract time
verb - physical action, active
verb - physical action, passive
verb - psychic action, active
verb - psychic action, passive
verb - physical action, "other"
verb - psychic action, "other"
adjective - descriptive, of measure
adjective - descriptive, of quality
adverb - simple "how"
adverb - simple "when"
adverb - simple "where"
adverb - simple "why"
adverb - clausal "how"
adverb - clausal "when"
adverb - clausal "where"
adverb - clausal "why"

The following information accompanied each word on these lists. All grammatical information applied to it by the analyst and the same grammatical information about the word immediately preceding and the word immediately succeeding (except where terminal punctuation immediately preceded or followed: in such cases only the word "period" was printed out). The sequence number of the word in the sample. The number of the sentence in which the word occurred and all information supplied for that sentence. The frequency of a given word on the lists had to be obtained by counting the number of occurrences printed out.

Some high-frequency words (taken from Thorndike's count, see Bibliography) were excluded from these lists, since inclusion of all words made for incredibly bulky print-outs. These "excluded" words were listed and the number of times they occurred in the sample indicated, but no further information about them was supplied by the program.

RESULTS

The amount of data compiled by this project is so enormous that the thorough analysis of it will take several years. Indeed, I shall here report only minimally on vocabulary findings, because this portion of the data requires specially painstaking consideration in the performance of meaningful evaluations. And even the record of syntactic analyses should be regarded as merely preliminary. I estimate that no more than 5 percent of the data which might provide valuable results under proper study has to date been really investigated. All my findings, futhermore, must be regarded as provisional, since, as I explain in the Discussion section, the total elimination of errors from the data is not yet feasible.

Note: All figures have been rounded for clarity of presentation, so some categories do not total exactly 100%. My aim here has not been absolute accuracy in detail but accurate representation of fundamental ranges, because the interest in all these figures is their value in comparisons or contrasts.

RESULTS I

TABLE 1.1
SENTENCE LENGTH

A.	Death of AGA		Spanish Gypsy	
Total sentences	80		62	
Total words	2113		1810	
Average words/sentence	26.41		29.35	
B.				
number "narrative" sents. followed by average words/ sentence for this type	33	27.52	44	28.11
"dialog" sentences with following words/sent av.	26	24.08	9	19.67
C.				
Length in words of each of first 25 sentences (d indicates dialog sentence, m="mixed," no mark indicates "narrative"), followed by difference in length from preceding sentence	11 -		57 -	
	11 0		20 37	
	23 12		54 34	
	25 2		19 35	
	24 1		7 12	
	27 3		6 1	
	22 5		7 1	
	20 2		41 34	
	22d2		42 1	
	24d2		73 31	
	24d0		37 36	
	25d1		60m23	
	24d1		42d18	
	21d3		14d28	
	64 43		27d13	
	51d13		26d1	
	36d15		9 d7	
	24d12		1 d8	
	13d11		20d19	
	30d17		5 d15	
	31d1		21d16	
	22d9		61 40	
	26d4		30 31	
	25d1		13 18	
	28d3		18 5	

Range of difference	AGA	Gypsy
0-9 words	17	6
10-19 words	6	8
20-29 words	0	2
30-39 words	0	7
40-49 words	1	1

	AGA	Gypsy
Approx. average difference in length between sucessive sentences	7 words	20 words

TABLE 1.2
SENTENCE STRUCTURE

A. Number and average length	AGA		Gypsy	
simple sentences	18	18.17	12	15.25
compound sentences	24	30.04	6	24.33
complex sentences	11	20.45	23	29.65
compound-complex sentences	21	34.76	13	48.38
fragment	2	9.00	4	13.75
fragment with clause(s)	4	23.00	4	23.75
B. Number of narrative sentences				
simple	5		10	
compound	6		5	
complex	7		19	
compound-complex	11		3	
fragment	1		3	
fragment with clause	3		4	
Total	33		44	
Number of dialog sentences				
simple	7		2	
compound	9		0	
complex	4		3.	
compound-complex	5		3	
fragment	1		1	
fragment with clause	0		0	
	26		9	
C. Percent of total clauses				
noun clauses	20.4		7.9	
adjective clauses	42.9		60.7	
adverb clauses	36.7		31.5	

TABLE 1.3
PARTS OF SPEECH

A. Parts of speech, percent total words	AGA	Gypsy
noun	20.8	23.9
pronoun	7.4	6.5
verb (simple)	8.0	7.9
verb (main part, compound)	4.4	3.6
verb (auxiliary, compound)	4.5	3.8
verbal	4.6	4.6
adjective	25.8	26.8
adverb	8.8	6.1
preposition	8.0	10.8
conjunction	7.3	5.2
contraction	0.5	0.1
interjection	0.1	0.4
B. Percent of four parts of speech (totals for all sentences in parentheses) in simple sentences		
noun	36.2 (31)	34.4 (35)
verb	14.5 (18)	15.6 (17)
adjective	36.7 (38)	40.0 (39)
adverb	12.6 (13)	10.0 (9)
compound-complex sentences		
noun	28.5 (31)	34.9 (35)
verb	20.3 (18)	15.6 (17)
adjective	36.6 (38)	40.0 (39)
adverb	14.6 (13)	9.5 (9)
narrative sentences		
noun	30.8	34.0
verb	16.3	16.9
adjective	39.2	38.6
adverb	13.7	10.4
dialog sentences		
noun	31.4	33.6
verb	19.9	22.4
adjective	38.6	39.3
adverb	10.1	4.7

TABLE 1.4
NOUNS

A. Number and percent of nouns identified as	AGA		Gypsy	
concrete - person	39	15.1	61	35.5
concrete - place	28	10.8	4	2.3
concrete - things	192	74.1	107	62.2
concrete - total	259		172	
abstract - quality	35	21.1	38	15.9
abstract - action	4	2.4	10	4.2
abstract - idea	25	15.1	49	20.5
abstract - collective	83	50.0	125	52.3
abstract - time	19	11.4	17	7.1
abstract - total	166		239	
B. Number and percent of nouns in that class that are subject of sentence or clause				
concrete	68	26.3	34	19.8
abstract	30	18.1	41	17.2

TABLE 1.5
VERBS

A. Percentages of verb tenses	AGA		Gypsy	
present	23.0		23.1	
past	49.4		49.5	
future	3.1		4.3	
present perfect	3.8		1.0	
past perfect	3.1		5.3	
present progressive	0.4		0.0	
past progressive	4.2		0.5	
present "modal"	10.7		15.9	
present perfect "modal"	2.3		0.5	
B. Number and percent of total number of verbs				
present tense-"narrative"	22	8.4	13	6.3
past tense-"narrative"	51	19.5	79	38.0
present tense-"dialog"	26	10.0	8	3.8
past tense-"dialog"	37	14.2	3	1.4

C. Number and ratio for verbs classified "physical," "psychic," and "other" according to present, future, or past tense. Ratio in the upper right hand corner of each block of four applies to the "tense" row, the ratio in the lower left corner of the block applies to the classification column, and the ratio in the lower right corner applies to the total number of verbs.

AGA						Gypsy						
Physical		Psychic		Other		Physical		Psychic		Other		
31	.348	4	.045	54	.607	11	.136	11	.136	59	.728	Present
.333	.118	.211	.015	.360	.206	.440	.053	.324	.053	.396	.284	
1	.125	2	.250	5	.625	0	.0	1	.111	8	.889	Future
.011	.004	.105	.008	.033	.019	.0	.0	.029	.005	.054	.038	
61	.370	13	.079	91	.552	14	.119	22	.186	82	.695	Past
.656	.233	.684	.050	.607	.347	.560	.067	.647	.106	.550	.394	

D. Percent of verbs	AGA	Gypsy
active	84.7	79.8
passive	4.6	5.8
copulative	10.7	14.4
transitive	37.8	47.1
intransitive	49.2	34.1
copulative	13.0	18.8
transitive in "narrative"	36.3	47.7
intransitive in "narrative"	54.9	34.6
transitive in "dialog"	48.8	75.0
intransitive in "dialog"	41.7	12.5

E. Percent of types of present tense	AGA	Gypsy
universal	1.9	60.4
regular	62.3	37.5
imperative	35.8	2.1

TABLE 1.6
ADJECTIVES

A. Percent adjectives classified as	AGA	Gypsy
descriptive of measure	2.2	5.4
descriptive of quality	27.9	39.6
definite article	16.7	18.1
indefinite article	6.8	11.1
demonstrative	5.7	4.5
numerical	1.1	0.6
pronominal	18.5	13.0
other - "definite"	16.7	5.2
other - "indefinite"	4.4	2.5

B. Number and ratio of adjectives immediately preceding noun classes

AGA						Gypsy						
Concrete		Proper		Abstract		Concrete		Proper		Abstract		-- NOUNS
0	.000	0	.000	4	1.00	0	.000	1	.091	10	.909	Descriptive Adj. of measure
.000	.000	.000	.000	.056	.019	.000	.000	.333	.005	.081	.047	
22	.595	1	.027	14	.378	15	.263	1	.018	41	.719	Descriptive Adj. of quality
.167	.107	.500	.005	.194	.068	.172	.070	.333	.005	.333	.192	
45	.804	0	.000	11	.196	44	.571	1	.013	32	.416	Articles
.341	.218	.000	.000	.153	.053	.506	.207	.333	.005	.260	.150	
49	.700	1	.014	20	.286	24	.471	0	.000	27	.529	demonstrative, numerical, pronom
.371	.238	.500	.005	.278	.097	.276	.113	.000	.000	.220	.127	
16	.410	0	.000	23	.590	4	.235	0	.000	13	.765	other
.12.	.078	.000	.000	.319	.112	.046	.019	.000	.000	.106	.061	

	AGA	Gypsy
Total Number concrete nouns	259	172
Concrete nouns with immed. preceding adjective	132	87
Total number abstract nouns	166	239
Abstract nouns with immed. preceding adjective	72	123

C. Number of descriptive adjectives separated by	0 words	1 word	2 words	3words	4 words
AGA	3	13	16	10	11
Gypsy	21	22	27	23	16

TABLE 1.7
ADVERBS

	AGA		Gypsy	
A. Number and percent of adverbs classified as				
Answering question "how?"	73	39.0	31	27.9
Answering question "when?"	28	15.0	13	11.7
Answering question "where?"	37	19.8	22	19.8
negative	23	12.3	21	18.9
expletive	13	7.0	18	16.2
intensifier	9	4.8	3	2.7
other	4	2.1	3	2.7
B. Number and percent of adverbs in degree				
positive	31	16.6	12	10.8
comparative	4	2.1	3	2.7
superlative	1	0.5	4	3.6
"degree" not applicable	151	80.7	92	82.9

TABLE 1.8
CONJUNCTIONS

A. Number and percent of conjunctions	AGA		Gypsy	
coordinate, linking clauses	38	24.5	17	17.9
coordinate, not linking clauses	81	52.3	48	50.5
correlative	3	1.9	0	0
subordinating	33	21.3	30	31.6
Total	155		95	
B. Number and percent of total number of conjunctions				
coordinating & correlative in narrative sentences	56	36.1	43	45.3
coordinating & correlative in dialog sentences	31	20.0	4	4.2
subordinating in narrative sentences	12	7.7	20	21.1
subordinating in dialog sentences	12	7.7	3	3.2

RESULTS 2
TABLE 2.1
PARTS OF SPEECH
PERCENT OF TOTAL WORDS

	Noun	Pronoun	Verb Simple	Verb Compound	Verbal	Adjective	Adverb	Preposition	Conjunction	Total Verb
Defoe: Moll Flanders	13.7	15.9	9.0	9.2	7.2	16.4	8.9	9.5	9.4	18.2
Richardson: Pamela	13.0	16.3	8.4	10.6	7.5	15.7	9.0	8.4	9.8	19.0
Fielding: Joseph Andrews #2	20.1	12.6	9.1	8.3	6.5	17.1	8.3	9.3	7.6	17.4
Joseph Andrews #1	19.0	12.2	8.0	9.2	5.8	19.8	6.4	10.4	8.1	17.2
Tom Jones	19.6	10.2	6.1	8.7	7.1	20.2	8.8	12.8	6.0	14.8
TOTAL										
Burney: Evelina	15.9	15.4	8.8	9.5	6.6	14.7	10.3	8.9	6.5	18.3
Scott: Old Mortality	21.9	9.2	6.0	7.7	7.0	20.8	6.1	14.3	6.9	13.7
Austen: Love and Friendship	20.7	9.5	6.2	6.8	7.9	20.4	6.8	13.6	7.7	13.0
Northanger Abbey #2	15.9	12.8	8.0	8.8	6.3	19.4	10.9	10.2	6.8	16.8
Lady Susan	16.8	12.9	7.8	9.1	7.2	17.4	9.7	11.0	8.1	16.9
Sense and Sensibility	18.9	9.7	7.4	9.2	7.0	18.9	9.7	11.5	6.5	16.2
Pride and Prejudice	18.7	10.2	7.8	9.5	6.2	18.2	10.2	11.1	7.6	17.3
Mansfield Park	18.6	10.2	6.0	8.0	7.1	20.4	8.1	11.8	8.3	14.0
Emma #2	17.0	11.9	7.8	11.2	6.8	18.9	10.2	9.0	6.4	19.0
Persuasion #2	19.6	10.2	7.0	10.9	6.5	19.7	7.6	10.5	7.5	17.9
#2 TOTAL										
Northanger Abbey #1	18.3	10.8	7.0	9.7	6.4	20.0	8.0	10.5	8.2	16.7
Emma #1	19.3	10.7	5.4	10.8	7.3	18.4	8.2	10.1	8.5	16.2
Persuasion #1	19.5	9.6	7.1	8.6	6.5	20.0	8.7	10.7	6.8	15.7
TOTAL										
Dickens: Oliver Twist	21.0	8.7	9.8	5.8	5.8	21.0	8.9	9.6	7.0	15.6
David Copperfield	16.2	14.6	10.5	8.2	6.4	16.4	9.4	9.2	7.4	18.7
Bleak House	18.4	11.1	8.7	6.0	7.0	18.7	9.1	10.5	8.3	14.7
Great Expectations	16.7	14.7	9.1	6.8	6.7	16.9	9.3	10.9	7.6	15.9
Our Mutual Friend	15.3	17.7	9.8	8.2	5.8	15.7	8.3	10.4	7.5	18.0
TOTAL										

TABLE 2.1--Continued

	Noun	Pro-noun	Verb Simple	Verb Compound	Verbal	Adjective	Adverb	Preposition	Conjunction	Total Verb
Thackeray: *Vanity Fair*	22.4	8.1	7.5	8.3	5.3	22.2	6.2	11.7	7.4	15.8
Charlotte Brontë: *The Spell*	18.1	11.7	9.0	7.8	6.2	20.6	8.6	9.5	8.0	16.8
Jane Eyre #2	17.9	14.9	11.0	8.8	5.4	17.7	8.8	8.3	6.7	19.8
Jane Eyre #1	18.0	13.4	10.0	6.6	5.2	19.7	7.9	8.5	7.1	16.6
The Professor	19.0	11.6	8.9	7.0	7.1	20.9	6.7	11.4	6.6	16.9
Shirley	21.3	12.6	9.4	7.5	4.4	21.2	7.0	9.4	6.9	16.9
Villette	18.4	12.0	8.8	7.8	6.7	20.4	7.0	9.6	6.7	16.6
Emily Bronte: *Wuthering Heights*	16.7	14.0	9.7	8.1	6.7	17.9	8.2	9.5	7.2	17.8
Anne Bronte: *Agnes Grey*	18.0	11.4	6.7	8.4	7.0	20.8	7.1	10.3	9.2	15.1
Tenant Wildfell Hall	15.6	15.6	7.4	8.2	6.1	16.2	9.1	10.1	9.1	15.6
TOTAL										
Trollope: *Way We Live Now*	19.3	13.1	7.6	11.4	5.3	15.9	7.2	10.0	7.0	19.0
Eliot: *Scenes Clerical Life*	21.3	6.2	6.8	7.0	7.3	23.0	7.6	11.8	8.4	13.8
Adam Bede	16.9	10.5	7.4	8.9	7.1	19.1	8.2	9.9	6.8	16.3
Mill on the Floss	18.8	11.6	8.2	7.9	6.9	17.9	7.6	9.9	7.4	16.1
Silas Marner	19.2	8.6	6.3	7.6	7.6	21.8	7.5	11.5	6.8	13.9
Romola	22.0	8.5	6.7	9.2	5.3	23.7	5.9	12.7	5.8	15.9
Felix Holt	17.5	11.1	8.6	10.6	6.1	19.4	9.4	8.4	7.8	19.2
Middlemarch #2	19.2	11.0	8.2	9.1	7.0	18.3	8.5	10.6	7.5	17.3
Middlemarch #1	19.7	10.3	7.6	7.8	6.2	21.1	6.7	11.3	6.9	15.4
Daniel Deronda	19.9	10.3	7.1	9.5	6.1	21.0	6.9	12.1	6.3	16.6
TOTAL										
Hardy: *Far From Madding Crowd*	15.9	14.0	8.7	7.8	6.6	17.8	9.9	9.8	6.7	16.5
Jude the Obscure	21.8	8.7	7.3	6.2	6.1	23.3	7.7	11.1	7.6	13.5
TOTAL										
Lawrence: *Sons and Lovers*	18.4	13.8	13.4	7.2	4.3	18.1	9.5	7.1	4.6	20.6
Woolf: *Jacob's Room*	24.6	8.4	9.9	5.1	5.2	21.1	6.7	10.5	7.6	15.0
Greene: *Power and Glory*	20.7	10.1	9.3	7.0	3.9	22.1	7.9	10.1	5.3	16.3
Joyce: *Portrait of the Artist*										

TABLE 2.2 VERBS
TENSE USE IN DIALOG AND NARRATIVE

Percent Verbs Simple Present / Percent Verbs Simple Past For	All Verbs	In Narrative Sentences	In Dialog Sentences
Defoe: *Moll Flanders*	22/50	4/73	58/0
Richardson: *Pamela*	25/38	22/44	31/46
Fielding: *Joseph Andrews* #2	27/48	2/75	58/18
Joseph Andrews #1			
Tom Jones	24/44	5/69	50/0
TOTAL			
Burney: *Evelina*	20/47	11/66	36/10
Scott: *Old Mortality*	26/43	5/60	46/18
Austen: *Love and Friendship*	15/55	6/64	50/14
Northanger Abbey #2	32/38	1/71	52/15
Lady Susan	37/31	32/35	40/33
Sense and Sensibility	13/59	1/74	41/18
Pride and Prejudice	17/52	0/78	41/12
Mansfield Park	21/41	8/54	39/22
Emma #2	28/34	4/65	44/12
Persuasion #2	20/40	5/54	42/8
#2 TOTAL			
Northanger Abbey #1			
Emma #1			
Persuasion #1			
TOTAL			
Dickens: Oliver Twist	17/61	3/68	44/21
David Copperfield	29/47	2/76	58/13
Bleak House	35/41	25/52	55/9
Great Expectations	14/60	4/70	47/29
Our Mutual Friend	38/34	2/81	55/11
TOTAL			

TABLE 2.2--Continued

Percent Verbs Simple Present / Percent Verbs Simple Past For	All Verbs	In Narrative Sentences	In Dialog Sentences
Thackeray: *Vanity Fair*	9/62	2/66	22/41
Charlotte Brontë: *The Spell*	30/41	2/77	48/17
Jane Eyre #2	25/51	2/74	57/16
Jane Eyre #1			
The Professor	27/49	13/64	47/16
Shirley	37/36	2/76	50/22
Villette	15/57	10/65	32/33
Emily Bronte: *Wuthering Heights*	19/52	5/73	42/15
Anne Bronte: *Agnes Grey*			
Tenant Wildfell Hall			
TOTAL			
Trollope: *Way We Live Now*	13/47	1/64	28/14
Eliot: *Scenes Clerical Life*	21/49	8/65	83/0
Adam Bede	31/33	3/66	48/11
Mill on the Floss	16/57	7/62	23/40
Silas Marner	14/53	1/77	41/19
Romola	11/53	1/62	35/17
Felix Holt	29/35	2/54	51/18
Middlemarch #2	23/40	9/53	40/12
Middlemarch #1			
Daniel Deronda	27/39	4/54	50/19
TOTAL			
Hardy: *Far From Madding Crowd*	38/33	7/70	58/7
Jude the Obscure	18/57	2/67	50/19
TOTAL			
Lawrence: *Sons and Lovers*	15/65	1/85	43/19
Woolf: *Jacob's Room*			
Greene: *Power and Glory*	7/64	2/80	24/11
Joyce: *Portrait of the Artist*	9/60	4/65	46/23

TABLE 2.3 VERBS
TRANSITIVE AND INTRANSITIVE

Percent Transitive/Intransitive	All Verbs	Verbs in Narrative Sentences	Verbs in Dialog Sentences
Defoe: *Moll Flanders*	44/30	50/25	58/17
Richardson: *Pamela*	54/26	53/28	48/29
Fielding: *Joseph Andrews* #2	52/30	62/23	49/19
Joseph Andrews #1	66/17		
Tom Jones	44/33	42/40	61/17
TOTAL			
Burney: *Evelina*	41/43	45/33	46/35
Scott:*Old Mortality*	50/33	48/33	42/40
Austen: *Love and Friendship*	43/32	46/31	41/27
Northanger Abbey #2	45/25	38/34	49/17
Lady Susan	44/29	45/31	60/27
Sense and Sensibility	33/42	33/42	43/29
Pride and Prejudice	39/35	34/39	44/26
Mansfield Park	41/24	40/24	38/22
Emma #2	35/32	50/23	58/6
Persuasion #2	24/39	27/36	25/39
#2 TOTAL			
Northanger Abbey #1	55/24		
Emma #1	52/26		
Persuasion #1	35/37		
TOTAL			
Dickens: Oliver Twist	22/56	28/54	32/41
David Copperfield	36/39	44/36	38/31
Bleak House	40/34	43/29	48/28
Great Expectations	44/39	44/36	47/33
Our Mutual Friend	42/35	39/54	47/25
TOTAL			

TABLE 2.3--Continued

Percent $\frac{\text{Transitive}}{\text{Intransitive}}$	All Verbs	Verbs in Narrative Sentences	Verbs in Dialog Sentences
Thackeray: Vanity Fair	43/36	40/39	46/29
Charlotte Brontë: The Spell	38/39	42/43	42/35
Jane Eyre #2	43/40	43/46	50/36
Jane Eyre #1	37/38		
The Professor	48/32	51/27	49/34
Shirley	43/27	38/38	45/23
Villette	48/31	46/32	44/32
Emily Bronte: Wuthering Heights	45/37	46/37	52/30
Anne Bronte: Agnes Grey	36/39	35/42	29/30
Tenant Wildfell Hall	48/31	44/33	55/23
TOTAL			
Trollope: Way We Live Now	45/34	40/34	60/29
Eliot: Scenes Clerical Life	37/35	45/33	42/42
Adam Bede	44/33	28/42	56/25
Mill on the Floss	56/27	47/36	61/21
Silas Marner	45/30	35/34	57/23
Romola	44/33	47/30	44/39
Felix Holt	41/33	44/34	44/27
Middlemarch #2	24/41	23/39	28/33
Middlemarch #1	35/37		
Daniel Deronda	41/32	42/31	46/24
TOTAL			
Hardy: Far From Madding Crowd	39/34	32/42	49/24
Jude the Obscure	37/34	40/31	51/31
TOTAL			
Lawrence: Sons and Lovers	33/49	39/41	37/42
Woolf: Jacob's Room	59/25		
Greene: Power and Glory	32/53	32/48	54/35
Joyce: Portrait of the Artist	34/56	29/63	54/31

TABLE 2.4
ACTIVE AND PASSIVE VERB

Percent of All Verbs	Passive	Copulative
Defoe: Moll Flanders	6	25
Richardson: Pamela	4	21
Fielding: Joseph Andrews #2	3	17
Joseph Andrews #1	8	17
Tom Jones	9	23
TOTAL		
Burney: Evelina	6	17
Scott: Old Mortality	7	19
Austen: Love and Friendship	9	22
Northanger Abbey #2	6	29
Lady Susan	6	24
Sense and Sensibility	6	24
Pride and Prejudice	5	22
Mansfield Park	6	34
Emma #2	11	33
Persuasion #2	8	33
#2 TOTAL		
Northanger Abbey #1	12	19
Emma #1	11	20
Persuasion #1	7	23
TOTAL		
Dickens: Oliver Twist	5	22
David Copperfield	2	24
Bleak House	4	25
Great Expectations	5	17
Our Mutual Friend	4	21
TOTAL		

TABLE 2.4--Continued

Percent of All Verbs	Passive	Copulative
Thackeray: Vanity Fair	8	21
Charlotte Brontë: The Spell	3	20
Jane Eyre #2	2	16
Jane Eyre #1	2	20
The Professor	5	20
Shirley	3	26
Villette	5	20
Emily Bronte: Wuthering Heights	3	18
Anne Bronte: Agnes Grey	6	20
Tenant Wildfell Hall	6	21
TOTAL		
Trollope: Way We Live Now	6	20
Eliot: Scenes Clerical Life	5	24
Adam Bede	8	24
Mill on the Floss	4	17
Silas Marner	10	24
Romola	6	20
Felix Holt	2	24
Middlemarch #2	4	27
Middlemarch #1	3	22
Daniel Deronda	3	23
TOTAL		
Hardy: Far From Madding Crowd	5	24
Jude the Obscure	6	25
TOTAL		
Lawrence: Sons and Lovers	3	16
Woolf: Jacob's Room	8	14
Greene: Power and Glory	2	12
Joyce: Portrait of the Artist	10	6

TABLE 2.5
SIMPLE VERBS PHYSICAL ACTION AND PSYCHIC ACTION

Percent One-Word Verbs	Physical	Psychic	Ratio
Defoe: Moll Flanders	27	11	.41
Richardson: Pamela	37	18	.49
Fielding: Joseph Andrews #2	35	14	.40
Joseph Andrews #1			
Tom Jones	40	29	.73
TOTAL			
Burney: Evelina	55	17	.31
Scott: Old Mortality	30	18	.60
Austen: Love and Friendship	43	22	.51
Northanger Abbey #2	20	26	1.3
Lady Susan	38	24	.63
Sense and Sensibility	36	12	.33
Pride and Prejudice	35	9	.26
Mansfield Park	28	12	.43
Emma #2	25	19	.76
Persuasion #2			
#2 TOTAL			
Northanger Abbey #1			
Emma #1			
Persuasion #1			
TOTAL			
Dickens: Oliver Twist	71	3	.04
David Copperfield	43	14	.33
Bleak House	55	16	.29
Great Expectations	55	21	.38
Our Mutual Friend	44	11	.25
TOTAL			

TABLE 2.5--Continued

Percent One-Word Verbs	Physical	Psychic	Ratio
Thackeray: Vanity Fair	46	12	.26
Charlotte Brontë: The Spell	51	22	.43
Jane Eyre #2	41	20	.49
Jane Eyre #1			
The Professor	53	20	.38
Shirley	37	12	.33
Villette	64	8	.13
Emily Bronte: Wuthering Heights	54	21	.39
Anne Bronte: Agnes Grey	55	13	.24
Tenant Wildfell Hall	47	21	.45
TOTAL			
Trollope: Way We Live Now	44	13	.30
Eliot: Scenes Clerical Life	36	19	.53
Adam Bede			
Mill on the Floss	50	14	.28
Silas Marner	42	21	.50
Romola	23	20	.87
Felix Holt	33	13	.39
Middlemarch #2	12	20	1.7
Middlemarch #1			
Daniel Deronda	29	18	.62
TOTAL			
Hardy: Far From Madding Crowd	30	15	.50
Jude the Obscure	32	12	.38
TOTAL			
Lawrence: Sons and Lovers	65	6	.09
Woolf: Jacob's Room			
Greene: Power and Glory	70	8	.11
Joyce: Portrait of the Artist	62	13	.21

TABLE 2.6
DESCRIPTIVE AND LIMITING ADJECTIVES IMMEDIATELY PRECEDING NOUNS

	Total Number Descriptive Adjectives	Number Preceding Nouns	Percent	Total No. Limiting Adjectives	Number Preceding Nouns	Percent
Defoe: *Moll Flanders*	234	40	18	237	163	69
Richardson: *Pamela*	310	82	26	269	166	62
Fielding: *Joseph Andrews* #2						
Joseph Andrews #1						
Tom Jones						
TOTAL	489	138	28	382	283	74
Burney: *Evelina*						
Scott: *Old Mortality*	232	72	31	133	86	65
Austen: *Love and Friendship*						
Northanger Abbey #2						
Lady Susan						
Sense and Sensibility						
Pride and Prejudice						
Mansfield Park						
Emma #2						
Persuasion #2						
#2 TOTAL	1827	430	24	1112	729	65
Northanger Abbey #1						
Emma #1						
Persuasion #1						
TOTAL						
Dickens: Oliver Twist						
David Copperfield						
Bleak House						
Great Expectations						
Our Mutual Friend						
TOTAL	1467	310	21	1105	705	64

TABLE 2.6--Continued

	Total Number Descriptive Adjectives	Number Preceding Nouns	Percent	Total No. Limiting Adjectives	Number Preceding Nouns	Percent
Thackeray: Vanity Fair	418	89	21	314	187	60
Charlotte Brontë: The Spell						
Jane Eyre #2						
Jane Eyre #1						
The Professor						
Shirley						
Villette						
TOTAL	1135	264	23	769	534	70
Emily Bronte: Wuthering Heights	264	45	17	316	217	69
Anne Bronte: Agnes Grey						
Tenant Wildfell Hall						
TOTAL	476	154	32	472	307	65
Trollope: Way We Live Now	244	59	24	154	111	72
Eliot: Scenes Clerical Life						
Adam Bede						
Mill on the Floss						
Silas Marner						
Romola						
Felix Holt						
Middlemarch #2						
Middlemarch #1						
Daniel Deronda						
TOTAL	2536	606	24	1358	847	63
Hardy: Far From Madding Crowd						
Jude the Obscure						
TOTAL						
Lawrence: Sons and Lovers	110	50	45	199	138	70
Woolf: Jacob's Room						
Greene: Power and Glory	293	28	10	55	39	71
Joyce: Portrait of the Artist						

TABLE 2.7
ADJECTIVES, DESCRIPTIVE, LIMITING, ARTICLES, DEGREES

Percent of all Adjectives	Descriptive	Limiting	Positive	Comparative	Superlative	Percent Total Words Definite Article	Percent Total Words Indefinite Article
Defoe: Moll Flanders	25.5	25.8	192	6	11	22.3	11.3
Richardson: Pamela	38.9	33.8	175	8	8	17.4	7.5
Fielding: Joseph Andrews #2	22.2	28.7	104	7	10	22.3	17.6
Joseph Andrews #1							
Tom Jones	38.1	23.3	181	13	6	26.8	10.7
TOTAL							
Burney: Evelina	32.6	34.1	179	0	13	20.7	11.8
Scott: Old Mortality	34.1	19.5	109	13	5	37.0	8.1
Austen: Love and Friendship	30.1	29.1	84	4	11	18.8	9.5
Northanger Abbey #2	28.3	20.9	178	7	8	15.9	9.9
Lady Susan	33.6	26.6	122	6	9	18.1	10.7
Sense and Sensibility	34.6	22.0	206	8	11	19.7	9.3
Pride and Prejudice	26.9	22.7	156	13	10	23.0	11.6
Mansfield Park	44.9	17.9	228	11	11	22.6	12.1
Emma #2	43.4	21.1	98	6	7	22.8	10.7
Persuasion #2	42.9	26.1	111	28	9	17.2	10.7
#2 TOTAL							
Northanger Abbey #1							
Emma #1							
Persuasion #1							
TOTAL							
Dickens: Oliver Twist	33.5	24.3	227	26	15	31.2	10.2
David Copperfield	37.1	25.1	304	23	5	16.5	12.4
Bleak House	35.9	26.1	160	7	5	25.2	10.8
Great Expectations	31.1	31.6	216	16	10	24.1	10.5
Our Mutual Friend	41.0	23.1	101	1	2	22.1	11.7
TOTAL							

TABLE 2.7--Continued

Percent of all Adjectives	Descriptive	Limiting	Positive	Comparative	Superlative	Percent Total Words: Definite Article	Percent Total Words: Indefinite Article
Thackeray: Vanity Fair	30.9	23.2	214	14	7	32.9	8.1
Charlotte Brontë: The Spell	35.9	26.6	122	1	6	21.2	10.9
Jane Eyre #2	33.3	30.8	128	11	3	18.7	11.7
Jane Eyre #1							
The Professor	32.4	27.5	122	7	6	24.6	13.9
Shirley	43.1	24.8	197	5	3	18.8	10.3
Villette	41.7	21.5	232	10	6	24.0	11.9
TOTAL							
Emily Bronte: Wuthering Heights	30.0	35.4	172	17	7	20.3	12.3
Anne Bronte: Agnes Grey	38.1	29.9	177	22	13	17.6	12.4
Tenant Wildfell Hall	32.0	38.1	164	21	3	17.1	11.4
TOTAL							
Trollope: Way We Live Now	36.8	23.2	135	6	4	28.2	19.4
Eliot: Scenes Clerical Life	37.5	18.5	225	10	14	23.3	10.7
Adam Bede	43.2	19.3	242	34	22	21.9	11.1
Mill on the Floss	37.6	25.7	180	12	7	21.6	11.2
Silas Marner	40.0	19.2	294	14	11	22.4	11.5
Romola	34.8	18.0	135	5	1	27.5	11.1
Felix Holt	36.8	20.9	186	20	9	15.7	12.4
Middlemarch #2	38.0	24.2	123	12	5	22.1	11.3
Middlemarch #1							
Daniel Deronda	36.6	18.5	314	12	12	21.0	12.3
TOTAL							
Hardy: Far From Madding Crowd	31.2	23.7	124	8	2	20.5	12.2
Jude the Obscure	30.2	15.6	122	5	3	25.8	15.3
TOTAL							
Lawrence: Sons and Lovers	36.7	24.2	185	7	2	18.0	12.4
Woolf: Jacob's Room							
Greene: Power and Glory	23.0	11.5	71	6	2	40.0	16.3
Joyce: Portrait of the Artist							

TABLE 2.8
VERBS - INFINITIVES AND PRESENT PARTICIPLES

Percent Total Words	Infinitives	Present Participles
Defoe: *Moll Flanders*	2.8	.04
Richardson: *Pamela*	2.7	.06
Fielding: *Joseph Andrews* #2	2.6	1.2
Joseph Andrews #1		
Tom Jones	2.1	.09
TOTAL		
Burney: *Evelina*	2.3	1.2
Scott: *Old Mortality*	2.3	.08
Austen: *Love and Friendship*	2.9	1.1
Northanger Abbey #2	2.5	.05
Lady Susan	2.1	.06
Sense and Sensibility	2.4	.06
Pride and Prejudice	2.3	.05
Mansfield Park	2.9	1.0
Emma #2	2.5	.06
Persuasion #2	2.5	.05
#2 TOTAL		
Northanger Abbey #1		
Emma #1		
Persuasion #1		
TOTAL		
Dickens: Oliver Twist	1.7	1.3
David Copperfield	2.4	1.1
Bleak House	2.4	1.6
Great Expectations	2.3	1.1
Our Mutual Friend	2.0	1.2
TOTAL		

TABLE 2.8--Continued

Percent Total Words	Infinitives	Present Participles
Thackeray: Vanity Fair	1.7	.08
Charlotte Brontë: The Spell	1.7	1.6
Jane Eyre #2	2.1	.07
Jane Eyre #1		
The Professor	2.1	1.4
Shirley	1.8	.07
Villette	2.1	1.7
TOTAL		
Emily Bronte: Wuthering Heights	2.1	1.3
Anne Bronte: Agnes Grey	2.5	.09
Tenant Wildfell Hall	2.2	.08
TOTAL		
Trollope: Way We Live Now	2.0	.09
Eliot: Scenes Clerical Life	2.0	1.0
Adam Bede	2.5	.06
Mill on the Floss	2.5	1.2
Silas Marner	2.2	1.2
Romola	1.8	.06
Felix Holt	2.2	1.0
Middlemarch #2	2.5	.09
Middlemarch #1		
Daniel Deronda	2.0	.08
TOTAL		
Hardy: Far From Madding Crowd	2.5	1.1
Jude the Obscure	1.9	1.0
TOTAL		
Lawrence: Sons and Lovers	1.7	1.5
Woolf: Jacob's Room		
Greene: Power and Glory	1.3	1.7
Joyce: Portrait of the Artist		

TABLE 2.9
VERB TENSES:
PROGRESSIVE AND MODAL

Percent of Total Separate Tenses	Progressive Present	Progressive Past	Modal Present	Modal Past
Defoe: Moll Flanders	1.7	1.5	12.8	1.8
Richardson: Pamela	.03	.04	10.3	5.3
Fielding: Joseph Andrews #2	.02	.06	11.9	1.3
Joseph Andrews #1				
Tom Jones	.02	.02	10.6	3.7
TOTAL				
Burney: Evelina	3.0	2.7	8.2	6.6
Scott: Old Mortality	1.9	.09	2.8	6.9
Austen: Love and Friendship	2.5	4.3	5.6	2.5
Northanger Abbey #2	4.3	3.1	10.4	.02
Lady Susan	2.6	.03	13.0	1.0
Sense and Sensibility	.02	.07	12.9	1.1
Pride and Prejudice	3.7	3.7	11.0	3.5
Mansfield Park	2.4	2.4	8.5	10.7
Emma #2	.03	1.0	7.2	13.4
Persuasio [illegible]	.08	3.9	16.3	2.0
#2 TOTAL				
Northanger Abbey #1				
Emma #1				
Persuasion #1				
TOTAL				
Dickens: Oliver Twist	1.6	3.1	5.2	1.0
David Copperfield	1.2	1.1	10.5	1.9
Bleak House	5.4	2.9	4.6	3.7
Great Expectations	2.6	2.0	5.6	3.9
Our Mutual Friend	3.5	1.4	11.8	7.6
TOTAL				

TABLE 2.9--Continued

Percent of Total Separate Tenses	Present	Past	Present	Past
	Progressive		Modal	
Thackeray: Vanity Fair	.07	3.3	2.2	8.2
Charlotte Brontë: The Spell	2.5	1.3	9.6	5.4
Jane Eyre #2	1.0	.00	11.5	1.8
Jane Eyre #1				
The Professor	1.3	2.8	3.5	3.5
Shirley	4.2	3.1	9.2	.06
Villette	2.6	4.5	5.3	3.6
TOTAL				
Emily Bronte: Wuthering Heights	2.9	3.1	5.3	5.5
Anne Bronte: Agnes Grey	.03	1.5	7.9	17.3
Tenant Wildfell Hall	5.0	3.3	7.8	9.5
TOTAL				
Trollope: Way We Live Now	3.4	5.5	13.8	1.8
Eliot: Scenes Clerical Life	2.3	5.3	9.2	2.7
Adam Bede	2.2	4.1	7.4	3.5
Mill on the Floss	1.7	2.6	5.1	6.5
Silas Marner	3.0	3.5	8.7	5.8
Romola	.08	1.9	10.2	4.9
Felix Holt	2.4	3.4	11.1	1.5
Middlemarch #2	4.3	3.0	10.8	1.8
Middlemarch #1				
Daniel Deronda	1.0	2.3	13.7	3.7
TOTAL				
Hardy: Far From Madding Crowd	5.5	2.7	7.9	1.2
Jude the Obscure	1.3	1.7	11.2	1.7
TOTAL				
Lawrence: Sons and Lovers	3.9	3.4	3.8	3.7
Woolf: Jacob's Room				
Greene: Power and Glory	2.8	8.0	9.7	.07
Joyce: Portrait of the Artist				

TABLE 3.1
MULTIPLE-WORD VERBS AND ONE-WORD VERBS, SUMMARIZED

	.28
Defoe: *Moll Flanders*	.61
Richardson: *Pamela*	
Fielding: *Joseph Andrews* #2	
Joseph Andrews #1	
Tom Jones	.55
TOTAL	.55
Burney: *Evelina*	.63
Scott: *Old Mortality*	
Austen: *Love and Friendship*	
Northanger Abbey	
Lady Susan	
Sense and Sensibility	
Pride and Prejudice	
Mansfield Park	
Emma #2	
Persuasion #2	.63
#2 TOTAL	
Northanger Abbey #1	
Emma #1	
Persuasion #1	
TOTAL	
Dickens: Cliver Twist	
David Copperfield	
Bleak House	
Great Expectations	
Our Mutual Friend	.31
TOTAL	

TABLE 3.1--Continued

Thackeray: Vanity Fair	
Charlotte Bronte: The Spell	
Jane Eyre #2	
Jane Eyre #1	
The Professor	
Shirley	
Villette	.37
TOTAL	.43
Emily Bronte: Wuthering Heights	
Anne Bronte: Agnes Grey	
Tenant Wildfell Hall	.61
TOTAL	
Trollope: Way We Live Now	
Eliot: Scenes Clerical Life	
Adam Bede	
Mill on the Floss	
Silas Marner	
Romola	
Felix Holt	
Middlemarch #2	
Middlemarch #1	
Daniel Deronda	.56
TOTAL	
Hardy: Far From Madding Crowd	
Jude the Obscure	.70
TOTAL	.26
Lawrence: Sons and Lovers	.27
Woolf: Jacob's Room	.44
Greene: Power and Glory	
Joyce: Portrait of the Artist	

TABLE 3.2
RATIO OF DESCRIPTIVE ADJECTIVES TO VERBS
(INCLUDING INFINITIVES AND PRESENT PARTICIPLES)

Defoe: Moll Flanders	.25
Richardson: Pamela	.35
Fielding: Joseph Andrews #2	
Joseph Andrews #1	
Tom Jones	
TOTAL	.38
Burney: Evelina	.27
Scott: Old Mortality	.52
Austen: Love and Friendship	.46
Northanger Abbey #2	.46
Lady Susan	.37
Sense and Sensibility	.43
Pride and Prejudice	.32
Mansfield Park	.49
Emma #2	.50
Persuasion #2	.55
#1 TOTAL	.44
Northanger Abbey #1	
Emma #1	
Persuasion #1	
TOTAL	
Dickens: Oliver Twist	.46
David Copperfield	.33
Bleak House	.43
Great Expectations	.31
Our Mutual Friend	.36
TOTAL	.37

TABLE 3.2--Continued

Thackeray: Vanity Fair	
Charlotte Brontë: The Spell	.42
Jane Eyre #2	.32
Jane Eyre #1	.40
The Professor	.48
Shirley	.48
Villette	
TOTAL	.44
Emily Bronte: Wuthering Heights	.29
Anne Bronte: Agnes Grey	
Tenant Wildfell Hall	
TOTAL	.41
Trollope: Way We Live Now	
Eliot: Scenes Clerical Life	.66
Adam Bede	.53
Mill on the Floss	.40
Silas Marner	.58
Romola	.60
Felix Holt	.42
Middlemarch #2	.42
Middlemarch #1	
Daniel Derona	.53
TOTAL	.51
Hardy: Far From Madding Crowd	
Jude the Obscure	
TOTAL	.41
Lawrence: Sons and Lovers	.33
Woolf: Jacob's Room	
Greene: Power and Glory	.31
Joyce: Portrait of the Artist	

TABLE 3.3
CONCRETE AND ABSTRACT NOUNS, SUMMARIZED

Percent of Concrete and Abstract Nouns	Concrete	Percent Total Nouns Concrete and Proper
Defoe: Moll Flanders	50	50
Richardson: Pamela	48	52
Fielding: Joseph Andrews #2		
Joseph Andrews #1		
Tom Jones		
TOTAL	53	59
Burney: Evelina	42	54
Scott: Old Mortality	37	46
Austen: Love and Friedship		
Northanger Abbey #2		
Lady Susan		
Sense and Sensibility		
Pride and Prejudice		
Mansfield Park		
Emma #2		
Persuasion #2		
#2 TOTAL	37	57
Northanger Abbey #1		
Emma #1		
Persuasion #1		
TOTAL		
Dickens: Oliver Twist		
David Copperfield		
Bleak House		
Great Expectations		
Our Mutual Friend		
TOTAL	52	60

TABLE 3.3--Continued

Percent of Concrete and Abstract Nouns	Concrete	Percent Total Nouns Concrete and Proper
Thackeray: Vanity Fair	57	65
Charlotte Brontë: The Spell		
Jane Eyre #2		
Jane Eyre #1		
The Professor		
Shirley		
Villette		
TOTAL	48	54
Emily Bronte: Wuthering Heights	44	49
Anne Bronte: Agnes Grey		
Tenant Wildfell Hall		
TOTAL	42	46
Trollope: Way We Live Now	53	64
Eliot: Scenes Clerical Life		
Adam Bede		
Mill on the Floss		
Silas Marner		
Romola		
Felix Holt		
Middlemarch #2		
Middlemarch #1		
Daniel Deronda		
TOTAL	44	50
Hardy: Far From Madding Crowd		
Jude the Obscure		
TOTAL	37	44
Lawrence: Sons and Lovers	73	80
Woolf: Jacob's Room	62	72
Greene: Power and Glory	72	75
Joyce: Portrait of the Artist	60	65

TABLE 3.4
PERSONS AS SUBJECTS OF SENTENCES AND CLAUSES

	Percent of all Subjects either Proper Names or Concrete Persons	Percent Total Nouns Serving as Subjects
Defoe: *Moll Flanders*	57	13
Richardson: *Pamela*	46	9
Fielding: *Joseph Andrews* #2		
Joseph Andrews #1		
Tom Jones		
TOTAL	70	12
Burney: *Evelina*		
Scott: *Old Mortality*	51	25
Austen: *Love and Friendship*		
Northanger Abbey #2		
Lady Susan		
Sense and Sensibility		
Pride and Prejudice		
Mansfield Park		
Emma #2		
Persuasion #2		
#2 TOTAL	58	19
Northanger Abbey #1		
Emma #1		
Persuasion #1		
TOTAL		
Dickens: Oliver Twist		
David Copperfield		
Bleak House		
Great Expectations		
Our Mutual Friend		
TOTAL	52	17

TABLE 3.4--Continued

	Percent of all Subjects either Proper Names or Concrete Persons	Percent Total Nouns Serving as Subjects
Thackeray: Vanity Fair	60	21
Charlotte Brontë: The Spell		
Jane Eyre #2		
Jane Eyre #1		
The Professor		
Shirley		
Villette		
TOTAL	36	17
Emily Bronte: Wuthering Heights	41	18
Anne Bronte: Agnes Grey		
Tenant Wildfell Hall		
TOTAL	52	13
Trollope: Way We Live Now	70	21
Eliot: Scenes Clerical Life		
Adam Bede		
Mill on the Floss		
Silas Marner		
Romola		
Felix Holt		
Middlemarch #2		
Middlemarch #1		
Daniel Deronda		
TOTAL	63	20
Hardy: Far From Madding Crowd		
Jude the Obscure		
TOTAL	40	20
Lawrence: Sons and Lovers	72	31
Woolf: Jacob's Room		
Greene: Power and Glory	35	28
Joyce: Portrait of the Artist		

PRECEDING PAGE MISSING

TABLE 3.6
COORDINATION AND SUBORDINATING CONJUNCTIONS NOT LINKING CLAUSES

Number	Narrative Coord.	Narrative Sub.	Dialog Coord.	Dialog Sub.	Narrative Ratio Coordinate to Subordinate	Dialog Ratio Coordinate to Subordinate
Defoe: Moll Flanders	75	97	1	3		
Richardson: Pamela	110	76	7	7		
Fielding: Joseph Andrews #2	38	30	6	16		
Joseph Andrews #1						
Tom Jones	28	16	2	5		
TOTAL	66	46	8	21	1.4	.38
Burney: Eveline	38	38	19	11		
Scott: Old Mortality	50	23	44	27		
Austen: Love and Friendship	35	34	7	6		
Northanger Abbey #2	38	33	36	35		
Lady Susan	41	52	0	9		
Sense and Sensibility	76	59	15	14		
Pride and Prejudice	41	61	16	40		
Mansfield Park	106	52	39	32		
Emma #2	28	20	22	22		
Persuasion #2	20	30	7	10		
#2 TOTAL	385	341	142	168	1.1	.85
Northanger Abbey #1						
Emma #1						
Persuasion #1						
TOTAL						
Dickens: Oliver Twist	38	47	9	11		
David Copperfield	79	92	42	49		
Bleak House	93	165	11	14		
Great Expectations	132	100	18	17		
Our Mutual Friend	26	13	35	44		
TOTAL	368	417	115	135	.88	.85

TABLE 3.6--Continued

Number	Narrative Coord.	Narrative Sub.	Dialog Coord.	Dialog Sub.	Narrative Ratio Coordinate to Subordinate	Dialog Ratio Coordinate to Subordinate
Thackeray: Vanity Fair	142	105	12	10		
Charlotte Brontë: The Spell	13	12	22	23		
Jane Eyre #2	34	28	26	22		
Jane Eyre #1						
The Professor	44	33	16	8		
Shirley	27	6	41	36		
Villette	88	26	21	10		
TOTAL	206	105	126	99	2.0	1.3
Emily Bronte: Wuthering Heights	85	56	27	29	1.5	.93
Anne Bronte: Agnes Grey	75	35	8	7		
Tenant Wildfell Hall	117	51	64	50		
TOTAL	192	86	72	57	2.2	1.3
Trollope: Way We Live Now	48	57	11	27		
Eliot: Scenes Clerical Life	53	39	2	1		
Adam Bede	33	24	58	41		
Mill on the Floss	74	63	48	52		
Silas Marner	48	24	18	33		
Romola	31	34	4	5		
Felix Holt	25	38	21	37		
Middlemarch #2	32	49	13	14		
Middlemarch #1						
Daniel Deronda	49	52	24	28		
TOTAL	345	323	184	211	1.1	.87
Hardy: Far From Madding Crowd	25	31	17	32		
Jude the Obscure	37	49	6	7		
TOTAL	62	80	25	39	.78	.64
Lawrence: Sons and Lovers	59	29	19	9		
Woolf: Jacob's Room						
Greene: Power and Glory	29	15	3	4		
Joyce: Portrait of the Artist	21	20	1	0		

TABLE 4.1
SENTENCE LENGTH

	Number Sentences	Average Length	Average Length Narrative Sentences	Average Length Dialog Sentences	Average Length Mixed Sentences	Ratio Dialog/Mixed	Dialog/Narrative	Mixed/Narrative	Number Fragmentary Sentences
Defoe: Moll Flanders	140	40.1	45.5	17.8	29.4	.61	.39	.65	3
Richardson: Pamela	186	27.3	26.4	16.6	26.9	.62	.61	1.00	9
Fielding: Joseph Andrews #2									7
Joseph Andrews #1									4
Tom Jones									0
TOTAL	380	32.6	41.0	19.0	31.0	.61	.46	.76	
Burney: Evelina	206	24.1	27.4	15.7	26.0	.61	.57	.95	15
Scott: Old Mortality	97	33.9	38.6	30.7	29.9	.98	.78	.78	10
Austen: Love and Friendship									10
Northanger Abbey #2									15
Lady Susan									2
Sense and Sensibility									8
Pride and Prejudice									4
Mansfield Park									13
Emma #2									10
Persuasion #2									4
#2 TOTAL									
Northanger Abbey #1									3
Emma #1									2
Persuasion #1									5
TOTAL	1736	23.8	31.2	15.5	26.0	.60	.50	.84	
Dickens: Oliver Twist									18
David Copperfield									16
Bleak House									15
Great Expectations									18
Our Mutual Friend									15
TOTAL	1162	18.2	26.8	10.9	18.5	.59	.41	.69	

TABLE 4.1--Continued

	Number Sentences	Average Length	Average Length Narrative Sentences	Average Length Dialog Sentences	Average Length Mixed Sentences	Ratio Dialog/Mixed	Dialog/Narrative	Mixed/Narrative	Number Fragmentary Sentences
Thackeray: Vanity Fair	258	23.6	25.1	16.2	22.6	.72	.65	.90	7
Charlotte Brontë: The Spell									17
Jane Eyre #2									5
Jane Eyre #1									6
The Professor									5
Shirley									6
Villette									9
TOTAL	1154	18.9	23.0	14.5	22.8	.64	.63	.99	
Emily Bronte:Wuthering Heights	275	18.1	26.7	10.4	16.0	.65	.39	.60	7
Anne Bronte: Agnes Grey									7
Tenant Wildfell Hall									1
TOTAL	231	29.8	38.2	22.2	28.3	.79	.58	.74	
Trollope: Way We Live Now	291	14.4	17.0	9.9	12.2	.81	.58	.72	9
Eliot: Scenes Clerical Life									7
Adam Bede									3
Mill on the Floss									8
Silas Marner									3
Romola									0
Felix Holt									4
Middlemarch #2									9
Middlemarch #1									2
Daniel Deronda	1349	24.1	31.6	17.6	23.4	.75	.56	.74	9
TOTAL									
Hardy: Far From Madding Crowd									12
Jude the Obscure									9
TOTAL	248	20.3	25.2	13.8	15.8	.87	.55	.63	
Lawrence: Sons and Lovers	439	10.0	10.7	8.2	10.1	.81	.77	.95	13
Woolf: Jacob's Room	561	16.1	16.1	7.8	12.9	.60	.48	.80	6
Greene: Power and Glory	136	15.9	20.7	7.6	12.7	.60	.38	.61	15
Joyce: Portrait of the Artist	180	14.2	14.5	8.5	14.7	.58	.59	.99	

TABLE 4.2
PROPORTIONS OF SENTENCE TYPES

Number of Sentences Classified as	Narrative	Dialog	Mixed	Ratio		
				Dialog / Mixed	Dialog / Narrative	Mixed / Narrative
Defoe: *Moll Flanders*						
Richardson: *Pamela*						
Fielding: *Joseph Andrews* #2						
Joseph Andrews #1						
Tom Jones						
TOTAL	167	82	136	.60	.49	.82
Burney: *Evelina*						
Scott: *Old Mortality*						
Austen: *Love and Friendship*						
Northanger Abbey #2						
Lady Susan						
Sense and Sensibility						
Pride and Prejudice						
Mansfield Park						
Emma #2						
Persuasion #2						
#2 TOTAL	796	760	167	4.5	.96	.21
Northanger Abbey #1						
Emma #1						
Persuasion #1						
TOTAL						
Dickens: Oliver Twist						
David Copperfield						
Bleak House						
Great Expectations						
Our Mutual Friend						
TOTAL	415	552	115	4.8	1.3	.28

TABLE 4.2--Continued

Number of Sentences Classified as	Narrative	Dialog	Mixed	Ratio		
				Dialog/Mixed	Dialog/Narrative	Mixed/Narrative
Thackeray: *Vanity Fair*						
Charlotte Brontë: *The Spell*						
Jane Eyre #2						
Jane Eyre #1						
The Professor						
Shirley						
Villette						
TOTAL	487	552	115	4.8	1.1	.24
Emily Bronte: *Wuthering Heights*						
Anne Bronte: *Agnes Grey*						
Tenant Wildfell Hall						
TOTAL						
Trollope: *Way We Live Now*						
Eliot: *Scenes Clerical Life*						
Adam Bede						
Mill on the Floss						
Silas Marner						
Romola						
Felix Holt						
Middlemarch #2						
Middlemarch #1						
Daniel Deronda						
TOTAL	529	575	245	2.3	1.1	.46
Hardy: *Far From Madding Crowd*						
Jude the Obscure						
TOTAL						
Lawrence: *Sons and Lovers*	207	85	126	.67	.41	.61
Woolf: *Jacob's Room*						
Greene: *Power and Glory*						
Joyce: *Portrait of the Artist*						

TABLE 4.3
SENTENCE STRUCTURE
TOTALS: SUMMARIZED

Percent Sentences Classified as	Simple	Compound	Complex	Comp-cx	Ratio		
					$\frac{S}{Cp-Cx}$	$\frac{S}{Cx}$	$\frac{S}{Cx+Cp-Cx}$
Defoe: Moll Flanders							
Richardson: Pamela							
Fielding: Joseph Andrews #2							
Joseph Andrews #1							
Tom Jones							
TOTAL	10	11	31	48	.21	.34	.13
Burney: Evelina							
Scott: Old Mortality							
Austen: Love and Friendship							
Northanger Abbey #2							
Lady Susan							
Sense and Sensibility							
Pride and Prejudice							
Mansfield Park							
Emma #2							
Persuasion #2							
#2 TOTAL							
Northanger Abbey #1							
Emma #1							
Persuasion #1							
TOTAL	30	13	28	29	1.1	1.1	.44
Dickens: Oliver Twist							
David Copperfield							
Bleak House							
Great Expectations							
Our Mutual Friend							
TOTAL	29	20	29	22	1.3	1.0	.57

TABLE 4.3--Continued

Percent Sentences Classified as	Simple	Compound	Complex	Comp-cx	Ratio		
					$\frac{S}{Cp\text{-}Cx}$	$\frac{S}{Cx}$	$\frac{S}{Cx+Cp\text{-}Cx}$
Thackeray: *Vanity Fair*							
Charlotte Brontë: *The Spell*							
Jane Eyre #2							
Jane Eyre #1							
The Professor							
Shirley							
Villette							
TOTAL	31	23	18	28	1.1	1.7	.67
Emily Bronte: *Wuthering Heights*							
Anne Bronte: *Agnes Grey*							
Tenant Wildfell Hall							
TOTAL							
Trollope: *Way We Live Now*							
Eliot: *Scenes Clerical Life*							
Adam Bede							
Mill on the Floss							
Silas Marner							
Romola							
Felix Holt							
Middlemarch #2							
Middlemarch #1							
Daniel Deronda							
TOTAL	23	14	30	32	.72	.78	.37
Hardy: *Far From Madding Crowd*							
Jude the Obscure							
TOTAL							
Lawrence: *Sons and Lovers*	50	19	13	5	10.0	4.0	2.9
Woolf: *Jacob's Room*							
Greene: *Power and Glory*							
Joyce: *Portrait of the Artist*							

TABLE 4.4
NARRATIVE AND DIALOG
SENTENCE STRUCTURE,
SUMMARIZED

	Percent Narrative Sentences				Ratios Narrative Sentences			Percent Dialog Sentences				Ratio Dialog Sentences		
	Simple	Compound	Complex	Cp-Cx	Simple/Cp-Cx	Simple/Complex	Simple/Cx+Cp-Cx	Simple	Compound	Complex	Cp-Cx	Simple/Cp-Cx	Simple/Complex	Simple/Cx+Cp-Cx
Defoe: Moll Flanders														
Richardson: Pamela														
Fielding: Joseph Andrews #2														
Joseph Andrews #1														
Tom Jones														
TOTAL	8	9	40	43	.20	.22	.10	29	13	32	26	1.1	.92	.50
Burney: Evelina														
Scott: Old Mortality														
Austen: Love and Friendship														
Northanger Abbey #2														
Lady Susan														
Sense and Sensibility														
Pride and Prejudice														
Mansfield Park														
Emma #2														
Persuasion #2														
#2 TOTAL														
Northanger Abbey #1														
Emma #1														
Persuasion #1														
TOTAL	25	11	31	33	.71	.82	.40	42	10	32	16	2.6	1.3	.88
Dickens: Oliver Twist														
David Copperfield														
Bleak House														
Great Expectations														
Our Mutual Friend														
TOTAL	25	12	41	22	1.1	.60	.49	52	11	27	10	4.9	1.9	1.4

TABLE 4.4--continued

	Percent Narrative Sentences				Ratios Narrative Sentences			Percent Dialog Sentences				Ratios Dialog Sentences		
	Simple	Compouns	Complex	Cp-Cx	Simple / Cp-Cx	Simple / Complex	Simple / Cx+Cp-Cx	Simple	Compound	Complex	Cp-Cx	Simple / Cp-Cx	Simple / Complex	Simple / Cx+Cp-Cx
Thackeray: Vanity Fair														
Charlotte Brontë: The Spell														
Jane Eyre #2														
Jane Eyre #1														
The Professor														
Shirley														
Villette														
TOTAL	30	24	20	26	1.1	1.5	.64	38	19	19	24	1.6	2.0	.87
Emily Bronte: Wuthering Heights														
Anne Bronte: Agnes Grey														
Tenant Wildfell Hall														
TOTAL														
Trollope: Way We Live Now														
Eliot: Scenes Clerical Life														
Adam Bede														
Mill on the Floss														
Silas Marner														
Romola														
Felix Holt														
Middlemarch #2														
Middlemarch #1														
Daniel Deronda														
TOTAL	21	6	40	33	.64	.52	.29	33	13	30	24	1.4	1.1	.61
Hardy: Far From Madding Crowd														
Jude the Obscure														
TOTAL														
Lawrence: Sons and Lovers	71	9	19	1	71.0	3.7	3.5	75	4	20	1	15.0	3.7	3.6
Woolf: Jacob's Room														
Greene: Power and Glory														
Joyce: Portrait of the Artist														

DISCUSSION

The results given in the preceding section are flawed by four potential kinds of error.

1. "Machine" errors. The widespread belief that a computer doesn't make mistakes is true only in the sense that a properly cared-for typewriter doesn't make mistakes. But computers must have operators, just as typewriters must have typists, and, like most well-used typewriters, computers are usually not in absolutely perfect working condition. Machine errors ordinarily are easily recognized but are expensive to correct. (I include in this category errors made by the programmer which usually have the same characteristics as actual machine errors.)

2. Keypunching errors. These "typos" of the computing world in my work, which now includes well over a quarter of a million records, cannot be totally eliminated in a short time. The computer itself can be used to correct some, but others give every appearance of being valid data and can be exposed only by painstaking re-analysis of all the data. The process of correction is in some respects more difficult than ordinary proofreading and is more time-consuming.

3. Analyzing errors. These errors occur when the analyst marks down the wrong code number for the information or puts the right number in the wrong column of the sheet he marks. A small number of these errors are undetectable, since some of the analyses depend on subjective decisions.

I was lucky in finding extremely careful and conscientious analysts and keypunchers, and the error percentage has been on the whole much lower than I had anticipated, but some errors of levels two and three have not yet been erased from the data. Too great a passion for error-free print-outs would be misguided at this time, first, because the statistical tabulations and averagings used are crude, but also -- and far more important -- because the fourth level of error, "textual" error, casts doubt on the accuracy of every part of the study.

I have used cheap, easily-available texts because I discovered very early that accurate texts at any price would not be available for some novels I wished to analyze. At the pioneering stage at which I have been working one can, I believe, grin and bear the problem of textual inaccuracy. But continued work along these lines will have to solve the problem of textual validity, and it will not be an easy one. Involved in it is not merely the matter of a "correct" printed version but the relationship of that version to what the novelist actually wrote. For example, we know that the first editions of George Eliot's novels are punctuated quite differently from the manuscripts she submitted. We have no way of knowing, however, how much of the punctuation of Jane Austen's novels represents the author's habits and how much her publisher's. At present one can, I repeat, legitimately disregard "textual inaccuracy," but even slightly more sophisticated analyses will require prolonged and careful attention to it.

For the reasons given in the paragraphs above, my findings at present are limited and provisional. It seems best, therefore, to present my evaluations in two parts.

1. An illustrative example of a detailed contrast which suggests the richness of the data and demonstrates the method by which ultimately it will be possible to make more-or-less definitive comparison-contrasts between the dictional and syntactic aspects of diverse novelists' styles.

2. A relatively succint explanation of my findings from the data relating directly to the specific objectives of the project based on the "total" output of the project.

In the section of this report entitled "Conclusions and Implications" I shall draw attention to aspects of broader significance adumbrated by these specific findings.

Discussed below are selected figures obtained for two poems, Emily Brontë's The Death of A.G.A. (abbreviated as AGA) and George Eliot's The Spanish Gypsy (abbreviated as Gypsy). Since these are poems, they ought not to be treated with the novels in this study, but they can serve to illustrate the methods of contrastive analysis applicable to novels made possible by this project. The material from the poems, moreover, is relevant in my view to any final description of their authors' fictional styles. At the risk of needless repetition, let me make this point clear.

I assume that a novelist's style is not a simple, rigidly limited set of characteristics but a developing process of shiftingly multi-inter-related characteristics. I am not so much interested in isolating idiosyncratic features by which one might identify the author of an anonymous piece of prose as I am in identifying patterns by which one might define the changing coherence of any novelist's development as a creator of fictions. This is why quite early in my work I decided to study relatively small samples (statistically too small) of a good many novels instead of larger samples from fewer novels. Ideally I ought to consider for contrastive definition non-fictional writings of the novelists, but funds so far have not permitted this experiment, except for the works reported on below. But the validity of my findings, ultimately, will depend on their extensiveness, the degree, that is, to which they take a meaningful place within a context defined by other kinds of writing and the degree to which they can be related systematically to the total corpus of other, associated writers' fiction.

TABLE 1.1

Discussion of Results.

The definition of "sentence" used in the study is the simplest graphological one: the words bounded by full stops. But of course even the nature of the "full stop" is open to debate (exclamation points and dashes have bothered analysts the most), so even the apparently plain matter of sentence length is in fact dubious. But gross differences in average sentence length do not seem to be of much significance. See the discussion later in this section. The material, incidentally, includes information as to the speaker of dialog, so that discriminations among the speeches of different characters can be established.

More valuable than gross average length of sentence are the various distribution patterns that can be worked out (the computer prints out one such graph). The simplest sort of distribution is illustrated in 1.1C, which makes it clear that AGA's sentences do not vary in length as much as Gypsy's.

TABLE 1.2

Perhaps the most striking feature of table 1.2A is the large percentage of compound sentences in AGA and the large percentage of complex sentences in *Gypsy*. Also noteworthy is the relative length of compound-complex sentences in *Gypsy* accompanying the relative brevity of simple sentences. Clause positions, though not indicated in the above tables, seem to me equally important, for example in AGA no sentences *begin* with a subordinate clause, while six of *Gypsy's* sixty-two so *commence*. It is intriguing, too, that 11 of AGA's thirty-three *narrative* sentences are compound-complex, more than half the total for *the entire* sample, whereas only three of Gypsy's forty-four narrative sentences are compound-complex, less than a quarter of the total. Dialog figures are obviously too small to be meaningful for *Gypsy*, but that nine out of twenty-six dialog sentences in AGA are *compound* emphasizes Emily Bronte's fondness for certain kinds of symetrical antithesis. Clause function figures in 2C point to *Gypsy's* tendency toward the adjectival and AGA's relatively heavier *emphasis* on the noun clause. Clause length, incidentally, shows variations, but on a scale small enough to make useful distinctions difficult: AGA's main clauses average 8.73 words in length (in *narrative* alone 8.93), subordinate clauses average 7.31 words in length (*narrative* alone = 7.46); for *Gypsy* the equivalent figures are main clauses = 9.68 (narrative alone = 10.97), subordinate clauses = 8.26 (narrative alone = 7.53).

TABLE 1.3

Most impressive is the basic similarity of the figures in 1.3A. Consistency in total percent of basic parts of speech is characteristic of all our print-outs (see below). Regularity in the percentage of each part of speech is still notable when break-down by sentence structure or type of discourse is applied, as in 1.3B and 1.3C, where only the slight rise in percent of nouns in AGA's simple sentences and the drop in percent of verbs in Gypsy's dialog sentences (probably the latter a result of the smallness of the sample) catches the eye. Results would seem to show conclusively that gross figures for parts of speech are not going to be good indicators of style. More subtle analysis of ratios of parts of speech may eventually be rewarding. There is the hint of a possible discrimination, for example, in the fact that in the figures for AGA, although the parts-of-speech percentages in simple and compound-complex sentences stay very close to the percentages for the total sample, they show contrary movements away from the norm provided by the total. Thus there is a higher percentage of nouns in simple sentences and a lower percentage in compound-complex sentences than the norm, while verbs in the two categories "move" in the opposite direction -- these movements not being found in Gypsy. While one would not want to base hypotheses, let alone draw conclusions, from the meagre figures presented, it does seem conceivable that such patterns of shiftings from a self-established norm might eventually be found to be stylistically important. But it will take some time to develop statistically valid results from the material.

TABLE 1.4

Any distinction between "abstract" and "concrete" nouns is an arbitrary one; furthermore, "abstractness" or "concreteness" is frequently determined by context, e.g., the word "clothing." This being true, the validity of sub-categories is even more dubious. The distinctions among concrete nouns are reasonably obvious, but probably a tri-partite division of abstract nouns which lumped together "quality" with "action" and "idea" with "collective" would be the most viable. Our distinctions have some utility, however, in vocabulary analysis, and it may be that the higher percentage of "abstract-quality" nouns in AGA and the higher percentage of "abstract-idea" nouns in Gypsy are useful indicators of stylistic tendencies. Surely the relatively high percentage of "concrete-person" nouns in Gypsy, which tends toward the more "abstract," is interesting. But in general it is the totals for the broad categories which appear most impressive. Notice that 1.4B, focused on noun function, emphasizes AGA's greater concreteness. This emphasis may be supported by the fact that only 13 percent of AGA's nouns are plural in number, as against 17.1 percent plural for Gypsy.

Note that in the table proper nouns (14 in AGA and 21 in Gypsy) are excluded.

TABLE 1.5

In 1.5A only those tenses where there was at least one occurrence in one of the poems are listed. All our studies show a similar dominance of simple present and simple past. Observe, however, that the past tense is relatively as frequent in AGA as in Gypsy, although 31 or AGA's 80 sentences contain dialog, where the present tense ordinarily prevails (probable statistical inadequacies of the samples are indicated, however, by 1.5B).

Although the distinction between verbs of "physical" action and "psychic" action (with "other" to collect all doubtful cases, which here, as in all our studies, are the majority) is at least as arbitrary as that between "abstract" and "concrete" nouns, the figures in 1.5C are striking enough to suggest that something more than hallucination by the analyst is involved. The relatively "objective" passive voice, as in all novels studied, appears most infrequently. Variations between transitive and intransitive verbs are difficult to evaluate. Distinctions between different kinds of present tense are sometimes remarkably clear cut, as in 1.5E, where AGA's high percentage of imperatives reflects the large amount of dialog in the sample, but also may point to the fashion in which Bronte characters tend to speak. The heavy use of the "universal" present in Gypsy is to be associated with the reflective cast of the sample passage, but certainly will not suprise anyone familiar with George Eliot's novels.

TABLE 1.6

Adjective classification has turned out to be more difficult than anticipated, and the last two categories in table 6A are really catch-alls for items (e.g., proper nouns used as adjectives) not covered by other categories. Of course subjective judgment plays a strong role in adjective classification, too. But the relatively high percentage of descriptive adjectives of quality in Gypsy seems important. 1.6B gives some indication of adjectival positions, and these appear to be significant, but of course such figures must be used in conjunction with figures for the nouns modified. Descriptive adjective clustering, though more rewardingly examined through vocabulary lists, is shown clearly in 1.6C to be greater in Gypsy than AGA. There appears to have been very little work done on the specific problem of positional frequency illustrated by 1.6C, which strikes me as potentially intriguing. The data for my study is laid out in such a way that from it positional distributions of a number of grammatical and semantic classifications could be plotted with relative ease.

TABLE 1.7

Our categorizations of adverbs, again, is arbitrary: our "how" category includes "how much?" and seems to absorb most adverbs that would answer the question "why?" In 1.7A it appears that Gypsy's relatively heavy use of "negatives" and "expletives," seemingly at the cost of "how" adverbs, is more than a chance phenomenon. Although figures for adverbial degrees, 1.7B, are quite small here (as thoughout our study), they may have some value. Although to only 17 percent of Gypsy's adverbs can degree distinctions be applied (compared to 19 percent for AGA), of these over a third are in the comparative or superlative: AGA plainly is more "positive." Increased sample size is obviously needed here, as it is for definition of adverbial function, the significance of which is suggested by the fact that AGA shows only 7 clausal adverbs out of 187, while Gypsy shows 12 out of a total of 111.

TABLE 1.8

When I began this study I thought conjunctions would be a significant feature of it. Conjunction figures, however, have been disappointing in that what they indicate is often as easily represented by other figures. In 1.8A, for example, Gypsy's relatively high percent of subordinating conjunctions and AGA's relatively high percent of coordinating conjunctions merely substantiates figures for compound and complex sentences. Discriminations between use of conjunctions in narrative and dialog may be worthwhile, as is suggested by the figures for AGA in 1.8B. I have come to the conclusion that a difficulty with statistical analyses of so-called "function" words is that their meanings are so various and often so subtly shaded. This subtle variation is probably most obvious in prepositions, but presumably more significant in conjunctions. In the sentence "I took off my clothes and went swimming" the "and" means something quite different from the "and" of "The knives and forks are in the lower drawer." Whether one sees a distinction between the "because" in "Her eyes were red because she had been crying" and the "because" in "He climbed the mountain because it was there" depends probably upon one's point of view, but surely the student of fictional style ought to be alert to such nuances.

VOCABULARY LIST I

The following nouns occur at least four times in the respective works.

AGA: love 14, eye 11, heaven 8, day 6, heart 6, pain 6, earth 6 (ground 2, dust 1), life 5, sun 5, sky 5, death 4 (dead 3), tears 4, blood 4, night 4, gore 4, cheeks 4, hands 4, agony 4 (misery, sorrow 1, despair 1, anguish 1, woe 2).

Gypsy: will 8, love 8, life 6, man 6, voice 6, song 6, noble 5, presence 5.

Note that there are 18 nouns in this list of highest frequency for AGA and only 8 for Gypsy: the tendency of AGA toward repetition is observable in all parts of speech studied. But a weaknesses of frequency counts is suggested by the fact that one of the two words common to both lists, "love," stands at the top of both lists, but its eight occurrences in Gypsy are scattered in seven different sentences, while its fourteen occurrences in AGA are concentrated in only eight sentences. This fact might suggest another form of repetition in AGA, but it is not characteristic: the eight occurrences of "heaven" are found in seven scattered sentences, the six occurrences of "pain" in six dispersed sentences, and so on.

For what it is worth, Gypsy's most frequent nouns, "love" and "will" occur in complex or compound-complex sentences, respectively, six and seven times. 11 of the occurrences of "love" in AGA are in the same kind of sentences, but only two occurrences of "heaven" are found there. I make these observations principally to call attention to the possibility of studying the superimposition of "grammatical" and "semantic" patterns which our layout of data makes quite feasible. There are, however, several statistical pitfalls in such work. It is obvious, for example, that, since compound-complex sentences are on the average longer than simple sentences, in a passage with an equal number of simple and compound-complex sentences there is a greater chance of any given word or occurrence being found in a compound-complex sentence than in a simple one. The more general significance of this point, which I virtually disregard in this report, is that there are in fact several different ways in which one can define the "size" of a given passage, and the validity of one's measurements is likely to depend on the appropriateness of the definition used.

High-frequency nouns in Gypsy are more often modified by a preceding adjective, as the following table suggests.

Novel-Word	Occurrences	Occurrences with preceding Adjective	Preceding Adjective descriptive
AGA-love	14	9	2
Gypsy-love	8	5	2
Gypsy-will	8	8	3
Gypsy-presence	5	5	3
AGA-heaven	8	4	2
AGA-pain	6	5	1
AGA-earth	6	3	0

The greater variety of nouns (fewer repetitions) in Gypsy may be illustrated by the lists for "concrete" nouns classed as "things." Of the total of 192 in this category in AGA there are only 98 different words, while of Gypsy's total of 107 in this category 75 are different. (ratios of .51 and .70 respectively).

Let us assume that the end of a sentence is a position of some rhetorical emphasis. Nouns immediately preceding a full stop, then, should be of importance. Study of this positional factor could take many directions, but I draw attention only to some obvious points. 44 sentences in AGA end with a noun, 55 percent of the total number of sentences, while 32 out of a possible 62, 51.6 per cent, so conclude in Gypsy. The pattern here is virtually identical. But seven of these "concluding" nouns in AGA have obvious reference to the natural word, while only one, "wings", in Gypsy might -- unless "universe" is to be counted. Likewise eight of the nouns ending sentences in AGA refer to pain and suffering, and only two in Gypsy fall into this class. Although there is the possibility of 32 occurrences, (there are two repetitions in this group from Gypsy, "song" and "court"; the three repetitions from AGA are "dead," "pride," "sea") of a common "ending" noun, there are in fact only three such occurrences, which appears, indeed, to be a fair indication of the general range of differentiation in the two vocabularies. Finally, one might observe that a higher percentage of the nouns concluding sentences in Gypsy exceed one syllable in length, 15 out of 32, as against 8 multi-syllable nouns out of 44 in AGA.

Actually figures for parts of speech other than nouns concluding sentences are more interesting than the noun figures.

Vocabulary List 3

Number and percent concluding sentences	AGA		Gypsy	
nouns	44	55	32	51.6
verbs	10	12.5	14	22.6
adverbs	16	20	5	8.1
adjectives	5	6.3	4	6.5
pronouns	5	6.3	6	9.7
prepositions	0	0	1	1.5

The substantive meaning of nouns in the lists can of course be studied in many ways. A difficulty, to my mind at least, is that sometimes it is the single use of a word, rather than its frequent occurrence, which counts heavily in the determination of style, yet one tends in working with word lists to concentrate on frequent occurrences. At the extreme limit, the omission of a word maybe important to a style -- to cite the obvious example, a study of the vocabulary of French classical drama which did not recognize the more-or-less codified restrictions of diction which operated in it would be ludicrous. There are further problems. 15 nouns in *Gypsy* identify persons by their occupation or profession, but there is only one such noun in AGA. Most of these nouns in *Gypsy*, however, are found in a single passage, and it seems to me that it is the occurrence of such a clustering, rather than the words themselves, that tells us most about the style of *Gypsy*. Nevertheless, even crude classifications can sometimes highlight significant subject-matter concentrations. 32 nouns not in the exclusion group classified as "abstract-quality" are listed for AGA, 35 for *Gypsy*. The following list indicates the words in this category used more than once.

Vocabulary List 4

AGA: agony 4, grief 2, pain 2, pride 3, youth 2, total = 13.
Gypsy: fulness 2, goodness 2, rage 2, resolve 2, secrecy 2, weakness 2, will 8, total = 20.

Here, as against the usual pattern, Eliot repeats more than Bronte. Eliot's poem is concerned with these "qualities," particularly in relation to "will," as Bronte is not. Moreover, Bronte's "qualities" seem almost to require direct, dramatic representation, while Eliot's imply a presentation of more internalized complexity.

The following complete listing as given by the computer of active verbs classified as "other," that is, not defined by the analyst as either "physical" or "psychic" action) indicates the difference in vocabulary between AGA and *Gypsy* -- notice only three verbs are common to both lists.

Vocabulary List 5

AGA: assist, bathed, bear, befell, betrays, bless 2, born 2, bowed, break, clears, clothed, combined, complain, darkened, darkens, decreed, delay, die, drain, drowned, dwelt, dwindled, flashed, gathering, greet, grew 3, grieve, guard, impelled, kindles, lies, lived, live, marked, melted, mocked, overflowed, overpassed, past, pleaded 2, plead, pray, prove, quenched, raise, scorned, seen, shines, showed, shut, slake, slumbered, snatched, spare, sustain, swear, waning, woke, wronged.
Gypsy: accept, breathed, bred, breed 2, buy, ceased, changed, chose 2, clung, comes, command, cross, dare, defeat, departs, despatch, disobey, divide, drank, failed, fed, finds, fixed, frame, frets, gathers, guard, helped, hurry, imaged, leaves, lived, master, penetrated, pictured, plead, pretend, quells, rains, resists, rounded, overturned, sate, saved, seems, serve, shakes, shows, shrank, sickened, sought, spare, tightened, urged, vanished, vanish, waits, waked, work.

The difference between the lists above is particularly striking because the *proportion* of verbs not in the exclusion group is comparable.

Vocabulary List 6

	AGA		Gypsy	
	Total	not in exclusion	Total	not in exclusion group
Active-physical verbs	93	53	25	14
Active-psychic	19	13	34	24
Active-other	150	64	149	61

In AGA verbs, like nouns, are more frequently repeated. In AGA 38 different verbs occur more than once, in *Gypsy* only 19. Modification of verbs, however, does not follow the pattern of modification of nouns.

Vocabulary List 7

Number of active verbs listed out (not in exclusion group)	AGA	Gypsy
	133	100
Active verbs with adverb immediately *preceding*	24 (18%)	10 (10%)
Active verbs with adverb immediately *following*	20 (15%)	12 (12%)
Preceding adverbs of classes "how," "when," "where"	13	3
Following adverbs "how" etc	18	8

Only four words appear in the listings for descriptive adjectives of measure in the two samples. AGA shows four occurrences of "many" and six of "all," *Gypsy* 17 occurrences of "all," one of "many," one of "whole," and two of "full." Of the 10 occurrences in AGA, three immediately precede a noun, and one immediately follows a noun, whereas 11 of *Gypsy's* 21 immediately precede a noun and six immediately follow. *Gypsy's* tendency toward more descrptive adjectives of measure appears to be supported by the frequency count of adjectives in the exclusion group.

Vocabulary List 8

AGA: good 1, little 1, fair 3, long 1, sweet 2, young 1, low 1.
Gypsy: little 2, new 2, better 1, high 2, more 1, sweet 2, kind 1, young 2, side 1, great 3.

Descriptive adjectives of quality again illustrate AGA's repetitions: out of 144 occurrences listed for AGA there are 114 unique adjectives, whereas the same figures for *Gypsy* are 181 and 161 (ratios of .79 and .89). Of these 161 different adjectives 16 are hyphenated, while only 5 of AGA's 114 are hyphenated.

Not only the tendency toward repetition in AGA but also the difference in kind of adjectives used in the two poems is suggested by the following list of descriptive adjectives of quality used more than once.

Vocabulary List 9

AGA: brief, bright, cold, crimson, dear 3, deep, dim, dreary, false, mortal 4, mountain 3, pure 3, sudden 3, summer, true 3, vain, warm 3, white, wild 5.
Gypsy: divine, fresh, hateful, human, idle, momentary, mystic, obstinate, passionate 4, poor, ready-shapen, strong 6, supreme.

The relatively higher frequency of clustered adjectives noted in a grammatical table previously takes on more significance when studied in the vocabulary lists. Here we find only one sequence of two descriptive adjectives of *quality* in AGA, but 15 in *Gypsy*. An association of this clustering with *Gypsy's* relatively limited repetition would seem reasonable.

More repetition and generally a greater reliance on adverbs is manifested by the following tables.

Vocabulary List 10

Adverbs occurring more than once that appear in exclusion group:
AGA: away 5, now 9, almost 2, below 2, much 2, once 2, really 2, as 4, never 4.
Gypsy: soon 2, when 4.

Vocabulary List 11

Adverbs occurring more than once *not* in exclusion group:
AGA: then 6, back 6, long 5, more 5, down 4, there 4, far 3, last 3, vain 3, again 2, full 2, still 2, today 2.
Gypsy: there 8, back 2, more 2, most 2, then 2.

In AGA adverbs, unlike adjectives, are more frequently clustered than in *Gypsy*, in which only four occurrences of sequential "how," "when," "where" adverbs are found, while there are 15 such sequences in AGA. Although there is so little dialog in the passage from *Gypsy* that figures for dialog-use are suspect, of the 51 adverbs printed out for *Gypsy*, only one occurs in a dialog sentence, while of AGA's similar total of 95, 15 appear in dialog sentences. Once more it is the potential method of measurement rather than the actual results to which I wish to draw attention.

One may suggest the difference in quality of the adverbs in the two poems simply by listing those printed out which end in "-ly."

Vocabulary List 12

AGA: bitterly, brightly, clearly, faintly, fondly, hoarsely, humbly, importantly, nearly, quietly, scarcely, softly, sternly, sweetly, vainly. Gypsy: dreamily, dumbly, hardly, lustrously, multitudinously, strangely, subtly, suddenly.

The foregoing is intended to illustrate how the data compiled in this project may be used. The following part of the discussion, which is concerned with the meaning of all figures compiled, will concentrate on syntax to the exclusion of vocabulary for the sake of brevity.

DISCUSSION 2.
TOTAL FIGURES

General consistency with marked irregularity within the pattern of consistency is most easily illustrated by the figures for the traditional PARTS OF SPEECH, Table 2.1. Often figures for samples from different authors are closer than for different samples from the same author, even from the same novel. This consistent-irregular pattern is further exemplified by discriminations within parts of speech. Verbs illustrate the point. Whether we examine TENSE USE IN DIALOG AND NARRATIVE, Table 2.2, TRANSITIVE AND INTRANSITIVE verbs similarly distinguished, Table 2.3, ACTIVE AND PASSIVE VERBS, Table 2.4, or, to take an entirely different kind of distinction, a thoroughly subjective one, the division of SIMPLE VERBS into those of PHYSICAL ACTION, PSYCHIC ACTION (with the remainder classifed as "OTHER"), Table 2.5, we find lack of significant system. It is true, I admit, that the ratio of "psychic" to "physical" verbs shows a distinct drop in the twentieth-century group, and that five of the seven samples from Eliot show ratios of .50 and larger, whereas none of Dickens' ratios go beyond .38, the Brontës never reach .50, while Austen, unusually, shows the widest range of fluctuation here, from .26 to 1.3. It is noteworthy, too, that if Love and Friendship and Lady Susan (both published only long after Austen's death, and both epistolary in form) are omitted, Austen's novels regularly show more than 50 percent of the verbs classified as "other," a feature that distinguishes her novels from Dickens', the Brontës', and Eliot's early novels. I would not deny, either, that Austen uses more passive verbs than Dickens, Charlotte Brontë, or George Eliot, averaging about eight percent to the others', respectively, four percent, three percent, and five percent approximately. An even more striking pattern emerges when one adds to the figures for the passive verbs those for copulative verbs. Austen runs a notably high proportion of copulative verbs; she averages approximately 34 percent passive and copulative verbs. Similar figures for other novelists show Dickens about 26 percent, Charlotte Brontë 24 percent, Eliot 28 percent, the eighteenth-century novelists as a group 26 percent, and the four twentieth-century novelists as a group only 18 percent. Nor is this latter patterning eccentric: the reader should notice throughout the figures a tendency for Austen to associate a little more closely with Eliot than with the Brontës, who tend to stay close to Dickens, and a tendency for the twentieth-century group to be markedly distinguished from the rest. On more than one occasion, moreover, the "mid-eighteenth century group" (in which I include Defoe along with Richardson and Fielding) is closer to mid-nineteenth-century novelists than to Austen, Scott, and Burney.

Even granting the significance of such patterns -- and I admit that Austen's relative preference for the passive and the copulative fits with an unobtrusiveness or transpicuousness of style definable in several different ways in her prose, which above all else tries not to call attention to itself -- I find the overwhelming weight of the evidence on the side of lack of patterning. Figures for DESCRIPTIVE AND LIMITING ADJECTIVES IMMEDIATELY PRECEDING NOUNS Table 2.6, support this view, even though one would think that the factor of position might carry us further than simple frequency counts. But perhaps even more striking is the scattering of the figures for the DEFINITE AND INDEFINITE ARTICLE, Table 2.7.

The negativeness of these general results must be emphasized to serve as a qualifying background for those discriminations which do point to possible systematized distinctions. Some more of these, to stay with categories already discussed for the moment, appear in the figures for INFINITIVES AND PRESENT PARTICIPLES, Table 2.8, where there is a clear difference between Austen and Dickens, and a more subtle but still observable difference between Austen's and Eliot's proportions. Figures for PROGRESSIVE TENSES AND MODALS, Table 2.7, indicate broad movements across the centuries, the novels apparently reflecting a rise in the use of progressives (a characteristic discussed by Curme, see bibliography) and a fall in the use of what is in fact the subjunctive in the language as a whole. There appears to be a distinction in the use of simple verbs as opposed to "compound" verbs (verb phrases), although the RATIO OF MULTIPLE-WORD VERBS TO ONE-WORD VERBS, Table 3.1, ought to be weighted in terms of past-present distinctions. Likewise figures for the ratios of LIMITING AND DESCRIPTIVE ADJECTIVES, Table 2.7, show some differences, Eliot's proportion of descriptive to limiting adjectives being higher than Dickens', the Brontës', or Austen's, although I am most impressed by the striking difference between the figures here for Austen's last three novels and her earlier works. In several tables the reader will notice that Mansfield Park, Emma, and Persuasion cluster rather distinctively -- the one even moderately satisfactory example I have found of novelistic development.

One can of course combine categories. For example, one can determine the RATIO OF DESCRIPTIVE ADJECTIVE TO VERBS, Table 3.2, where each multiple-word verb is counted as one and infinitives and participles are added in. Here, as with many single categories, it is difficult to define the significance of distinctions which are suggested. Austen's proportion of adjectives to verbs runs higher than Dickens', about equal to Charlotte Bronte's, and a bit lower than Eliot's -- but in what measure does this help to describe Jane Austen's style? Glancing over all the figures so far considered, we can observe the tendency of both Austen and Eliot to favor the adjective a bit more than Dickens or Charlotte Brontë (to notice only those authors for whom we have substantial totals) and the tendency of Eliot to favor descriptive adjectives more than limiting ones, in contrast to Austen at any rate. Austen also appears to be the only novelist studied who uses the superlative form of adjectives as much as the comparative, table 2.6, which characteristic, when one remembers that the overwhelming bulk of adjectival comparisons will occur with descriptive adjectives, further discriminates her from Eliot.

Yet these tendencies are so little pronounced that I at least am sure that the quality of the words employed, the semantic factor as opposed to the syntactic factor, is decisive in creating the particular effect of each author's style.

One basic problem in these analyses is to decide how to weight diverse elements. For example, the figures for PARTS OF SPEECH, Table 2.1, show clearly that noun and adjective percentages rise and fall together. One assumes that the adjectives "follow" the nouns. But it would be difficult to prove this assumption. It is worth pointing out, incidentally, that the consistency of the noun percentages is increased if one adds the pronoun figures to those for the noun alone.

The fuzzy and subjective distinction between ABSTRACT AND CONCRETE NOUNS, Table 3.3, produces some surprisingly interesting results. There is a notable movement toward increased concreteness in the twentieth-century authors. Jane Austen is the most consistently "abstract" of the authors extensively sampled (six different analysts working with 11 different samples arrived at very similar concrete-abstract ratios). Throughout, it may be observed, figures for Austen tend to be more consistent than for Dickens, the Brontës, and Eliot. It is interesting, too, that Burney and Scott (the latter almost an exact contemporary of Austen) are plainly more abstract, whereas in the middle of both the eighteenth and nineteenth centuries the figures hover closer to a 50-50 ratio. The pattern shifts strikingly, however, if PROPER nouns are added to the totals for concrete nouns, although the "concreteness" of the twentieth-century group is still emphatic. Austen, with one out of five nouns a proper noun, a name, becomes much more "concrete." Observe the figures for Defoe. In the passage from Moll Flanders sampled only one proper name occurred. This is an extreme case, but it is true, I believe, that Defoe uses relatively few names.

Interesting but puzzling are the differences in use of PERSONS OR THINGS AS SUBJECTS OF SENTENCES AND CLAUSES, Table 3.4, where "concrete nouns of person" and proper names are totalled against all other noun categories. These figures have to be judged along with those showing the percentage of all nouns serving as subjects, since there appears to be a trend from the eighteenth-century to the present toward an increase in this percentage. This trend is probably to be associated with the trend toward shorter sentences, but complex sentence structure can produce an equivalent effect.

For mysterious reasons the computer broke down more frequently when dealing with adverbs than with any other part of speech, and the figures for this category are incomplete. They are sufficient, however, to show some rather unexpected differences. Both Austen and Eliot appear to used fewer of what might be called "descriptive" ADVERBS (those answering the questions how, when, where, why) than the Brontës or Dickens, Table 3.5. More striking is Austen's reliance on INTENSIFIERS, particularly because this accompanies a high percentage of NEGATIVE adverbs. These characteristics appear to be identifiers of Jane Austen's fictional prose. Though not so clear as the foregoing, and with absolute numbers too small for statistical accuracy, figures for the DEGREE DISTINCTIONS AMONG ADVERBS are valuable -- note, for instance, the Brontës' stress on the positive degree and Austen's tendency toward a relatively high proportion of superlatives to positives. This tendency is in keeping with Austen's stylistic orientation toward "transparency" of language. She emphasizes with that part of speech, perhaps, where emphasis and "heightening" is least conspicuous. The kind of intensification she employs would be more blatant with adjectival forms. It seems to me characteristic of Austen to be emphatic unobtrusively.

Function words, while providing gratifyingly large numbers for statistical analysis, are difficult to interpret. PREPOSITIONS, for example, among PARTS OF SPEECH, Table 2.1, show a surprising range of variation. The eighteenth-century novels run toward low values here, Dickens' highs are near the bottom of Austen's range; Charlotte Brontë runs a trifle lower than Dickens, but with one sample going past all but Austen's three highest, and Eliot runs both high and low. The twentieth-century novels are low, except for Jacob's Room, which is quite high. Prepositions, in other words, re-emphasize that irregularity within general consistency which makes me dubious of the value of these syntactic statistics as stylistic descriptors of major importance.

CONJUNCTION figures, Table 2.1, are somewhat like those for prepositions, but, although the absolute figure totals are often too low to be worth much statistically, there do seem to be possible distinctions to be made between CONJUNCTION USE IN NARRATIVE AND DIALOG, Table 3.6, where one concentrates on conjunctions not used to link clauses. Eliot, Austen, and the Brontës favor such coordinating conjunctions in narrative sentences, though the Brontës' preference is stronger, and Eliot and Austen favor the subordinating conjunctions in dialog. The Brontës persist in their preference for coordinating conjunctions in dialog, while Dickens relatively consistently favors subordinating conjunctions in both narrative and dialog. Yet not only are the figures small, there are also shifts within the work of some of the authors -- Austen's samples vary markedly and Eliot's later novels are distinguishable from her earlier ones in this test. Here, as at many points, position, position both within the sentence and absolutely, that is, number of words apart regardless of full stops, is a dimension needed to help establish the significance of potentially interesting statistics.

The only elements other than parts of speech which it has been possible so far to study in any detail are sentence length and sentence structure. My figures on length in some respects run counter to the general belief that average sentence lengths in English prose have become progressively shorter. The existence of such a tendency (most thoroughly studied by Lucius Sherman) is deducible from my figures, but is at the least complicated by some counter-tendencies. My figures show clearly that in fictional prose a distinction must be made at least between "dialog" and "narrative" sentences, since all novelists studied reveal an average length for narrative sentences longer than for dialog sentences. Thorough analysis would distinguish among characters, too. To cite a relevant instance: Mr. Collins in Pride and Prejudice speaks fulsomely in rounded periods of empty complexity, whereas Mr. Woodhouse in Emma speaks brief, repetitive, infantile sentences. Jane Austen, I believe, likes to characterize by means of relatively subtle variations in sentence length and structure (with some characters, e.g. Miss Bates, the technique is not even subtle). Dickens, despite Alfred Jingle, seems to me to rely less on structural variation than on semantic peculiarity for individualizations, although structure frequently plays a role in his discriminations between social classes. But I have not yet had time to develop analyses of such discriminators. So far I have been able to deal only with broad differences between NARRATIVE (omitting description, exposition, soliloquy, etc), DIALOG, and MIXED NARRATIVE AND DIALOG.

Table 4.1, SENTENCE LENGTH, illustrates a significant means for developing one stylistic profile of fictional prose. Let me emphasize that I do not claim that the figures in the table provide such profiles. I would not be satisfied by figures from less than 3000 sentences per novel. It is the principle of establishing relationships between the major kinds of sentences (in another study I would lump all sentences not dialog or mixed into the narrative category) I wish to propose as a useful stylistic measure. The measure, of course, must be given its proper context. The first dimension of this is the basic absolute sentence length, that is, the average of all sentences combined. Thus Table 4.1 shows that the profile of ratios given in columns 6, 7, and 8 for Jane Austen and Virginia Woolf is virtually identical, but Austen's over-all sentence average is 23.8 to Woolf's 16.1. This illustration points up the fact that this measure simultaneously associates and discriminates, defines both likenesses and differences between authors, thus meeting the primary requisite for stylistic measurement.

A second contextual dimension is illustrated by PROPORTIONS OF SENTENCE TYPES, Table 4.2, (which uses only authors with samples of more than 300 narrative, dialog, and mixed sentences), where relationships between the number of sentences in each category are indicated. These provide one basis for judging the significance of the figures for average length. For example, the proportion of dialog sentences to narrative sentences is virtually identical in Dickens, Charlotte Brontë, and George Eliot, but the ratios of the lengths in these categories run from Dickens' .41 on an 18.2 base average sentence length through Eliot's .56 on a 24.1 base to Charlotte Brontë's .63 on an 18.9 base. Obviously, then, Dickens' characters' sentences are more discriminable, both relatively and absolutely, from narrative sentences than are the sentences of Brontë's characters' speeches. On this measure Eliot falls between Dickens and Brontë relatively but not absolutely (since her sentences are on the average longer). Austen's dialog is more differentiated from her narrative both absolutely and relatively than is Eliot's, though Austen, like Eliot, falls between Dickens and Brontë. Let me repeat, I do not claim that my figures accurately represent the facts, they are too small in number to do that. I am merely trying to illustrate a method of stylistic description that appears to me potentially useful, particularly if, instead of lumping all dialog together one discriminated among characters. Notice, too, that the figures in Table 4.2, permit the determination of relationships between sentence lengths and the distribution of sentence categories. Thus the profile of Fielding's sentence length relationships is reproduced almost identically by the profile of his numbers of sentences in each category -- a unique approach to repetition in this table.

One must insist, further, that the context for the sentence length profile is not complete without some reference to sentence structure. Although there is an obvious relationship between sentence length and sentence structure, a very short sentence is likely to be simple and a very long sentence is likely to be complicated, usefully discriminating definitions of sentence structure in fictional prose are not easily devised. First, of course, the distinctions between narrative, dialog, and mixed sentence have to be maintained, since the table for sentence length shows that average dialog sentences are shorter than average narrative sentences in every novel. It follows that where dialog predominates a tendency toward simpler structures should be manifest, and a reverse tendency where narrative predominates. Thus it could reasonably be argued that the striking differences between Lawrence and Fielding in SENTENCE STRUCTURE TOTALS, Table 4.3, are the product of Lawrence's heavier use of dialog and his much shorter sentence average. But Eliot and Austen, though running close on average sentence length and on proportion of dialog to narrative, diverge in Table 4.3. That we have here a structural, as opposed to a quantitative, discrimination is suggested by the fact that average length of both dialog and narrative sentences are reasonably close for the two authors. Interestingly, NARRATIVE AND DIALOG SENTENCE STRUCTURE, Table 4.4, shows structural variation between Austen and Eliot in both categories, with only a shade more differentiation in dialog than in narrative.

Even simpler indicators may be used, of course, as entrances into the establishment differentiations and likenesses: the relatively high proportion of compound sentences in Dickens and Charlotte Brontë strikes the eye in SENTENCE STRUCTURE TOTALS, Table 4.3. A careful reader may be puzzled to discover, however, that in NARRATIVE AND DIALOG SENTENCE STRUCTURE, Table 4.4, Dickens' percentages of compound sentences, unlike Brontë's, have dropped mysteriously. The fact is that an unusually high proportion of Dickens' mixed sentences (figures for which are not listed for the sake of clarity) are compound. Dickens is the only novelist I have studied who shows this liking for compound mixed sentences. Incidentally, it is only by attention to structural elements that one can distinguish significantly between Jane Austen's and Charlotte Brontë's dialog sentences. It is to me at least rather surprising how close the averages run, 15.5 to 14.5. Structurally, however, the differences are quite marked, with Brontë favoring compound and compound-complex patterns and Austen complex ones.

Some qualifications of these structural contrasts must be borne in mind. The traditional four categories I've used, while clear and understandable, are not necessarily the best that might be devised. They give no indication of phrases, for example. Nor do they indicate anything about position, yet a novel in which, say, all complex sentences began with subordinate clauses would be stylistically distinct from one in which all such sentences began with a main clause. Then there are non-sentence sentences: notice the final column in SENTENCE LENGTH, Table 4.1, which gives raw totals for fragmentary sentences, where Dickens stands out, but Austen scores surprisingly high. Finally, my figures indicate that very large samples are needed, probably at least 3000, since fluctuations are marked both in length and structure in all authors, and to obtain large figures for subdivisions a massive total is required.

Even giving full weight to these and other qualifications, one can regard this combination of profiles as a potentially useful stylistic measure, making it possible, for instance, to distinguish systematically between Jane Austen's style and George Eliot's, both of which have been described as "formal." Their narrative sentences are quantitatively equal, apparently, but Eliot's are more complex. Austen's dialog is quantitatively more distinct from her narrative than is Eliot's, and Austen's dialog tends to be structurally simpler than Eliot's. The complexity of the later novelist is emphasized by her preference for mixed sentences, which, though on the average shorter than Austen's, occur proportionately twice as frequently. The "formality" of Austen's style would seem to reside in a simplicity of structure in moderately long sentence units and a clarity of distinction between speech and narrative, whereas the "formality" of Eliot's resides in a careful articulation of element relationships within rather long sentence units. I confess, however, that far more interesting to me than any possible support for traditional descriptions of style is the simple discovery that Austen distinguishes so sharply between narrative and dialog, even to the point of using relatively few mixed sentences.

The question remains as to how much distinctions in sentence length and structure aid in the description of style. On one side my distinctions certainly need to be refined. Dialog should be studied, at the least, in terms of the differences between the characters who speak, as I suggested before. Discriminations among narrative sentences should be established, though not I believe along the lines I tried of "description," "exposition," and so forth. More purely formal discriminations would probably be more useful: classification by number of consecutive sentences uninterrupted by dialog, sentences involving action at, say, definable localities, with specific character groupings or in particular time sequences. In short what is needed is a simplified but relatively objective definition of some basic structures of the novel from which samples are derived. For an easy example, the significance of dialog sentence length's relation to narrative sentence length is to some degree dependent simply on the number of characters who speak and how much they speak, and these factors, in turn, relate directly to the size and complexity of the novel in which they appear. Thus the brevity and simplicity of Lawrence's sentences in _Sons and Lovers_ is most impressive when one contrasts these qualities with those of the sentences in _Silas Marner_ (70 percent complex and compound-complex, and 36.4 words average length), a much shorter novel with many fewer characters.

Or, to take a more speculative point: the _absolute_ length of

Scott's dialog (in proportion to narrative, notice, it scarcely differs from Lawrence's) can be related, I suspect, not only to the formal, public, "oratorical" situations in which his characters so often speak, but also to the simple fact that in a novel such as Old Mortality many of the characters speak relatively infrequently. This relative "silence" of Scott's characters is in sharp contrast to characters in Jane Austen's novels, who tend to "appear" only as "speakers." The "reality" of many of Austen's characters is created almost entirely by their speech. Yet Austen's dialog is not colloquial or "realistic." A good many of the Scottish characters in Scott's novels do speak with colloquial realism, and these attain a fictional "reality" I think quite different from that of most of Austen's characters, even though it is a "speaking" reality. But for others, like Burley, Macbriar, and Morton, three of the main speakers in my sample, "speech" serves a different function. It does not serve to establish the "reality" of the speaker as an individual character (it is no accident that Burley and Macbriar are literally "preachers") but to create the "reality" of the speaker's "historical" situation, that is, his special role in the dramatic unfolding of events encompassing a multitude of individuals.

CONCLUSIONS AND IMPLICATIONS

The preceding section makes it clear both explicitly and implicitly that the first seven specific objectives outlined for this project were not and could not be achieved. The fundamental and underlying objectives, however, were attained:

1: to compile an organized body of information, of a kind never before systematically collected, on the grammar and vocabulary of nineteenth-century novelists which might serve as a basis for further, more detailed, and more illuminating investigations of novelistic prose and for experiments in new techniques of teaching both literature and composition.

2: to lay a goundwork for systematic, relatively ofjective, and cumulatively-rewarding analyses of the nature and value of fiction.

So much more data, and so much richer data, than I had anticipated was collected by the project that I have not yet been able to analyze it fully. Indeed, it may be several years before such analysis is complete, and then it may well be possible to say that some, perhaps all, of the specific and limited objectives have been accomplished.

However, my principle conclusion from the work already done is that syntactic analysis is not an efficient, perhaps not even an adequate, method for describing novelistic style. I do not mean to suggest that there are not differences and patterns of difference in the syntax of the prose of different novels. My figures show that there are such differences. But I have grave doubts that these lead efficiently to better understanding and appreciation and definition of stylistic qualities. Even the sentence-length and sentence-structure tests I have proposed as stylistic indicators are, I believe, going to be of limited utility. If this is a negative conclusion, for me it is a pleasant one, because it implies that novelistic style is complex and operative on a level above that of regular language processes. If this were not so literature would not be worth studying; since it is so literature is worth studying -- and by methods not yet developed.

My work carries strong implications about the nature of some possible new methods that could be developed. First, I think my project shows that relatively systematized studies of literary phenomena are worth while. In the area of stylistic analysis my work indicates plainly that much larger bodies of data must be studied. We need bigger samples. To attain these in anything like the form I have used we must develop automated methods of grammatical analysis, vocabulary recognition, and the like. The basic work of this project was done manually. Such work is spiritually unrewarding and, if one seeks large masses of data, finally inefficient. It would be possible to develop computer programs capable of scanning texts (as an offshoot of my project we developed a primitive form of such a program), particularly if it is kept in mind that what is needed for literary study is not a program which will take into account all possible language events. On the contrary, what is needed is a "package" of small and relatively simple programs which will concentrate on specific, selected (and hopefully significant) items. For example, by using a dictionary of no more than 1000 "words" it would be possible, I estimate, to pick out automatically from any literary text about 80 percent of the syntactic material significant to stylistic studies. To put it simply, no literary scholar needs to know where every preposition is.

What is also required is more sophisticated statistical methods for the analysis of data. Some of these I hope to test against the data from this project in the next few years. Sophisticated statistical techniques also could be used to assist literary scholars in deciding what to search for, what to isolate, and what not to bother with -- before they begin.

Finally I think my work implies plainly that for the understanding and appreciation of literary style what is needed is an approach entirely different from that attempted in this project, one that does not reject the findings of syntactic and dictional analyses but complements them and adds to them an extra dimension of significance. This approach depends upon attacking prose style not from below, starting with units within the sentence, but from above, starting with fundamental unit of the work of art as a unified whole. The outline for such a procedure which I have developed is given in Appendix B. It has been applied to approximately twenty-five novels to date and results, so far as I have had time to study the material, are very impressive. The procedure I outline will certainly soon be modified and improved, by others, I hope, as well as myself. But the principle, that systematic, cumulatively-rewarding studies of literature can be devised which are purely literary, that is, not dependent on the theories and practices of linguistics, semantics, and other disciplines related but peripheral to literary criticism, has to my mind been firmly established by the results and implications of the work done in this project.

The foregoing brings up a point which has almost necessarily been slighted throughout this report: the relation of this work to the teaching of fiction. Through the cooperation of the English Department of the University of Wisconsin I was able last year to conduct a special class for fifteen graduate students planning to become teachers. The subject-matter of the course was provided by material collected by this project and related fictional material. The purpose of the course was to test new techniques for teaching novels and for interesting young people in the reading of novels. The course was highly successful in that all fifteen students found it enormously stimulating and intriguing, worked very hard, and were forced to think seriously about basic problems in the teaching and reading of fiction. It would be hard to define the specific results of this experiment, particularly since the class was far more successful at raising questions than at answering them, but we did arrive at a few definite conclusions.

Teaching of novels which involved consideration of stylistic and quantitative elements would almost certainly attract the interest of students ordinarily not responsive to literature.

A stylistic approach to the teaching of novels would surely force teachers to reconsider familiar subject-matter and would tend to revivify their enthusiasm for the material, since this approach does not require any rejection of old values or accepted classifications of fiction.

When only a few novels are studied, perhaps only two, by a given class, the stylistic method of approach might be the most rewarding, because it makes possible comparison-contrasts otherwise not attainable. An important corollary of this conclusion was that the new technique might make the establishing of relationships between novels of different periods (specifically, the relation of contemporary novels to "classics") not only easier but also much more meaningful.

The purpose of this experimental class was to investigate the teaching and reading of novels; we were not concerned with the relation of material from this project to composition teaching, and it has not yet been possible, as I had hoped, to experiment in this teaching area. But several of the students in the special class spontaneously raised the question of whether it might not be possible to adapt this kind of study of literature to some aspects of composition instruction. They and I saw the possibility that by making the *subject-matter* of a writing course the systematic analysis of writing one might attain a coherent, substantive content that is lacking from most classes in expository writing. This experiment would, in my opinion, be well worth trying.

SUMMARY

BACKGROUND

This project was an endeavor to define in a systematic and objective fashion fundamental characteristics of fictional prose style and to arrive at judgments of the nature and function of style in novelistic prose. Although this project has concentrated on the styles of particular nineteenth-century authors, on the relationships between the authors' styles, and on the development of fictional prose during the nineteenth century in Britain, the aim has been to establish principles applicable both to the study of literature generally, and, less directly, to the teaching of literature and composition. So little is known about the nature of literary style generally that the mere compilation of data resulting from this project provides a basis for modifying our present understanding of what any literary text consists of -- and this modification necessarily suggests new ideas about how literature should be taught. The material collected by this project, moreover, provides new insights into the nature of excellent prose and therefore has significant implications for the teaching of rhetoric, composition, and creative writing.

OBJECTIVES

The specific objectives of this study were summarily listed in the original proposal as follows:

1. To describe and to define the syntactic characteristics of the fictional prose style of five important nineteenth-century novelists (Jane Austen, Charlotte, Emily, and Anne Brontë, and George Eliot).

2. To describe and define the characteristic vocabulary found in the fictional prose of these five novelists.

3. To define changes and developments within each of the five authors' styles through study of the vocabulary and syntax of all the novels of each author.

4. To define special relationships which may exist between the styles of the three sisters Charlotte, Emily, and Anne Brontë.

5. To define relations between the style of Jane Austen, her novels having been published originally between 1811 and 1818, the Brontës, their novels having been published originally between 1847 and 1857, and George Eliot, whose novels originally were published between 1858 and 1876.

6. On the basis of the above to begin to define some factors characteristic of prose style in the nineteenth-century British novel.

7. To attempt to establish the relationship of dictional and syntactic characteristics of novelistic prose style to "macro-syntactic" characteristics, that is, features of plotting, characterization, and the like.

Beyond these specific aims, however, were the fundamental, underlying objectives of the project:

1. To compile an organized body of information, of a kind never before collected, on the grammar and vocabulary of nineteenth-century novelists which might serve as a basis for further, more detailed, and more illuminating investigations of novelistic prose and for experiments in new techniques of teaching both literature and composition.

2. To lay a groundwork for systematic, relatively objective, and cumulatively-rewarding analyses of the nature and value of fiction.

PROCEDURE

Sample passages as indicated below were selected from: Jane Austen, 8 novels; Charlotte Brontë, 5 novels; Emily Brontë, 1 novel; Anne Brontë, 2 novels; George Eliot, 8 novels; Charles Dickens, 5 novels; Henry Fielding, 2 novels; and at least one novel each by Daniel Defoe, Samuel Richardson, Fanny Burney, Walter Scott, Wm. M. Thackeray, Anthony Trollope, Thomas Hardy, James Joyce, D. H. Lawrence, Virginia Woolf, Grahame Greene.

Sample passages were of three kinds.

1. "Block" samples: all sentences from approximately ten consecutive pages, usually near the center of the novel. Ordinarily this would include at least one complete chapter.

2. "Random" samples: 10 to 50 units of five consecutive sentences, the units selected by line and the page by number by means of a table of random numbers.

3. "Special" samples: size and nature of these vary because they are selected to test a particular intuitive judgment or a specific hypothesis arising from study of other kinds of samples. Occasionally limited to a single kind of prose, e.g., only narrative, only dialogue, etc.

Sample passages in the novels indicated above were marked off. The text of the sample was typed on prepared sheets. Either I or an assistant marked these sheets, the marks providing grammatical or semantic information about each word, clause, and sentence. Also indicated were the characters present and the nature of their participation in the action. Furthermore, all relevant punctuational and paragraph information was coded. The completed information sheets were then given to a key punch operator who punched the data about each word and the word itself on a separate card. Card information was then transferred to magnetic tape. Computer work was conducted on a Control Data Corporation 3600 and miscellaneous card processing devices. Data cards, with one word of text per card plus information about the word and the sentence in which it occurs, were processed sequentially, providing a computer output of a listing of the total sample from which corrections could be made. Analytical procedures were divided into various sections, each corresponding to a definite problem, a specific desired set of information. At the completion of these procedures of statistical analysis and grouping the results were printed out. All these procedures are embodied in the programs written specifically for this project in Fortran.

RESULTS

1. The primary result of this study was the collection and organization of an enormous amount of carefully analyzed prose from the works of several important novelists. This material is now stored on magnetic tapes which can be easily reproduced. Students and scholars interested in style, in writing, in prose fiction, and in linguistics can obtain from records of this project data on syntactic patterns and vocabulary preferences which will be value to all kinds of studies other than that specifically undertaken in this project.

2. The second major result of this study was the establishment of some basic statistical data on the grammatical patterns of usage favored by five important nineteenth-century British novelists.

3. Another important result was the creation of frequency lists defining the favored vocabulary of five important nineteenth-century British novelists.

4. Finally, this study includes a significant amount of data of the kind listed in (2) and (3) immediately above from the works of twelve other major novelists. This data is interesting in itself and provides a framework within which to evaluate the information obtained about the five novelists chiefly studied, but, perhaps most important, it provides a basis for developing a systematized description of the novelistic style of an entire epoch.

CONCLUSIONS AND IMPLICATIONS

1. The primary conclusion of this study is that it is not possible to define the style of any novelist through simple statistical analysis of his grammar or of his word-choice. This conclusion, although a negative one, is of considerable importance for the study and teaching of literature. It could be stated positively in this fashion: style in sophisticated novels is an extremely complicated system of interrelationships between elements of many different kinds and levels, and style, therefore, cannot be explained or described adequately by methods which reduce or disregard its intrinsic complexity. My study appears

to show conclusively that grammar and vocabulary are only means by which a novelist achieves his style (as such, of course, they are important), so that any identification or characterization of a novelist's style stated solely in terms of grammatical patterns or vocabulary preferences found in his novels is inadequate. Mere authorship may be determined by a characteristic vocabulary or a preferred syntactic usage, but novelistic style can be satisfactorily identified only in terms of multiple factors, many of which go beyond the level of syntax and vocabulary. The chief proof of this lies in the fact that analyses of more than a quarter of a million words of novelistic texts in this study fail to provide definitive answers to any of the first six specific objectives of the project. The analyses do show, however, that there are quantitatively identifiable differences between the prose of different novelists.

2. Data from this study does seem to point, however, to the possibility of defining, albeit in broad and generalized terms, stylistic patterns of novelistic eras, e.g., eighteenth-century novel, twentieth-century novel, etc. Such definitions, however, to be accurate and valid will have to be founded upon analyses not simple but statistically rather sophisticated.

3. Deriving from (1) and (2) above is the conclusion that what is needed for the definition of style in fictional prose literature is the systematic analysis of factors that transcend sentence patterns and word-choice preferences (see specific objective 7 above). Included in the final report for this project is a statement of a procedure developed in the course of this study for such "macro-syntactic" analysis of whole novels.

4. The fundamental objectives of this study, the compilation of stylistic data never before available and the articulation of objective means for enriching our understanding of the nature of fiction, have been accomplished and point clearly to the value of further development of systematic investigations of literature.

5. Finally, two specific implications of this study point to the need for further work with the material collected during the course of this project. It appears likely that the development of more sophisticated statistical tests of the vocabularies and grammatic information collected would reveal much about the methods and purposes of specific novelists. Such tests, however, should be accompanied by further collection of data, since variations in analytical results of different parts of the same novel in this study have revealed that an adequate sample for the study of any novelist's style should probably contain about 60,000 consecutive words. The main implication of this conclusion, then, is that further systematic study of fictional prose style should be based on automated analysis of texts, not, as was true of this study, on human analyses of texts.

6. An experiment in teaching making use of material gathered for the project and the methods employed in its research was attempted. Results indicated that at least some of the techniques devised for the project would be useful in the development of more broadly interesting presentations of novelistic subject-matter in the classroom.

BIBLIOGRAPHY I
Computer Studies, Linguistics
Technical Analyses, etc.

Aurner, Robert R. "The History of Certain Aspects of the Structure of the English Sentence, "*Philological Quarterly*, II, 1923, 23-59

Barber, C. L. "Some Measurable Characteristics of Modern Scientific Prose," *Contributions to English Syntax and Philology*, Goteborg, 1962, 21-43

Bar-Hillel, Y. "The Mechanization of Literature Searching," *Proceedings of a Symposium on Mechanization of Thought Processes*, London, 1959, 789-800

Bar-Hillel, Y. "Theoretical Aspects of the Mechanization of Literature Searching," *Digital Information Processors*, ed. W. Hoffman, New York, 1962, 406-443

Blacklith, Robert. "A Multivariate Analysis of Latin Elegiac Verse," *Language and Speech*, VI, 1963, 196-205

Booth, Andrew D., L. Brandwood, and J. P. Cleave. *The Mechanical Resolution of Linguistic Problems*, London, 1958

Carroll, John B. "Vectors of Prose Style," *Style in Language*, ed. Thomas Sebeok, Cambridge, Mass, 1960, 283-292

Christensen, Francis. "A New Rhetoric," *College English*, XXV 1963, 7-11

Cherry, Colin. *On Human Communication*, New York, 1957

Chomsky, Noam. *Current Issues in Linguistic Theory*, The Hague, 1964

Chomsky, Noam. *Syntactic Structures*, The Hague, 1957

Curme, George O. *Syntax*, Boston, 1931

Dearing, Vinton A. *Methods of Textual Editing*, Los Angeles, 1962

Elderton, W. P. "A Few Statistics on the Length of English Words," *Journal of the Royal Statistical Society*, CXII, 1949, 436-443

Ellegard, Alvar. *A Statistical Method for Determining Authorship: The "Junius" Letters, 1769-1772*, Goteborg, 1962

Emma, Ronald D. *Milton's Grammar*, The Hague, 1964

Estrich, Robert M. and Hans Sperber. Three Keys to Language, New York, 1952

Fogle, Ephim G. "Electronic Computers and Elizabethan Texts," Studies in Bibliography, XV, 1962, 15-32

Garvin, Paul L., ed. Natural Language and the Computer, New York, 1963

Gerwig, George William, "On the Decrease in Predication of and of Sentence-Weight in English Prose," Univ. of Nebraska Studies, II, 1894, 17-24

Guiraud, Pierre. Problemes et Methodes de la Statistique, Dordrecht, 1959

Harris, Zellig S. String Analysis of Sentence Structure, The Hague, 1962

Hayes, D. G. Basic Principles and Technical Variations in Sentence Structure Determination, Santa Monica, California, 1960, p. 1984

Herdan, Gustav. Language as Choice and Chance, Groningen, 1956

Hildum, Donald C. "Semantic Analysis of Texts by Computer," Language, XXXIX, 1963, 649-653

Hruby, Antonin. "Statistical Methods in Textual Criticism," General Linguistics, V:3, 1962 (Supplement)

Lehman, W. P. "An Experiment in Machine Translation," The Graduate Journal, VII, 1965, 111-131

Milic, Louis T. "A Quantitative Approach to the Style of Jonathan Swift," Ph. D. Dissertation, Columbia University, 1963

Moritz, Robert E. "On Variation and Functional Relation of Certain Sentence Constants in Standard Literature," University of Nebraska Studies, III, 1903, 229-253

Mosteller, Frederich, and David L. Wallace. "Inference in an Authorship Problem," Journal of the American Statistical Association, LVIII, 1963, 275-309

O'Donnell, Bernard. "An Analysis of Prose Style to Determine Authorship: The O'Ruddy, a Novel by Stephen Crance and Robert Barr," Ph.D. Dissertation, Harvard University, 1963

Ohmann, Richard M. Shaw: The Style and the Man, Middletown, Conn, 1962

Parrish, S. M. "Problems in the Making of Computer Concordances," Studies in Bibliography, XV, 1962, 1-14

Perrin, Porter G. *Writer's Guide and Index to English* (Revised Edition), Chicago, 1950

Pike, Kenneth. "Language--Where Science and Poetry Meet," *College English*, XVI, 1965, 283-291

Riffaterre, Michael. "The Stylistic Function," *Proceedings of the 9th International Congress of Linguistics*, The Hague, 1964, 316-323

Riffaterre, Michael. "Fonctions du cliche dans la prose litteraire," CAIEF, no. 16, 1964, 81-95

Sebeok, Thomas A. "Notes on the Digital Computer as a Tool for Analyzing Literary Information," *Poetics*, ed. Donald Davie, 1961, 571-590

Sebeok, Thomas A. ed. *Style in Language*, Cambridge, Mass, 1960

Sedelow, Sally Y. and Walter a Sedelow, Jr and Terry Ruggles. "Some Paramaters for Computational Stylistics: Computer Aids to the Use of Traditional Categories of Stylistic Analysis," *Literary Data Processing Conference Proceedings*, New York, 1963, 211-229

Sherman, L. A. "Some Observations Upon the Sentence Lengths in English Prose" and "On Certain Facts and Principles in the Development of Form in Literature," in *University of Nebraska Studies*, I, 1888, 119-130 and I, 1892, 337-366

Spitzer, Leo. "Language--The Basis of Science, Philosophy and Poetry," *Studies in Intellectual History*, ed. George Boas et al., Baltimore, 1953

Tasman, Paul. "Literary Data Processing," *IBM Journal of Research and Development*, I, 1957, 249-256

Taube, Mortimer. *Computers and Common Sense*, New York and London, 1961

Thorndike, Edward L. and Irving Lorge. *The Teacher's Word Book of 30,000 Words*, New York, 1944

Williams, C. B. "Statistics as an Aid to Leterary Studies," *Science News*, XXIV, 1953, 99-106

Williams, C. B. "Studies in the History of Probability and Stylistics: A Note on an Early Statistical Study of Style," *Biometrika*, XLIII, 1956, 248-256

Yule, George. *The Statistical Study of Literary Vocabulary*, Cambridge, Eng., 1944

BIBLIOGRAPHY 2
Novel and Style Criticism
(Excluded in general are critical works on a single author)

Ackerman, J. S. "A Theory of Style," Journal of Aesthetics and Art Criticism, XX (1962), 227-237

Allen, Walter. The English Novel: A Short Critical History, New York, 1955

Allott, Miriam. Novelists on the Novel, London, 1959

Barish, Jonas A. Ben Jonson and the Language of Prose Comedy, Cambridge, Mass., 1960

Blandshard, Brand. On Philosophical Style, New York, 1954

Bond, H. L. The Literary Art of Edward Gibbon, Oxford, 1960

Booth, Wayne C. The Rhetoric of Fiction, Chicago, 1961

Boulton, Marjorie. The Anatomy of Prose, London, 1954

Brown, E. K. Rhythm in the Novel, Toronto, 1950

Cecil, David. The Early Victorian Novelists, Indianapolis, 1935

Chambers, R. W. On the Continuity of English Prose, Reprinted, New York, 1957

Cohen, Marcel. Grammaire et style, 1450-1950, Paris, 1954

Cook, Albert. The Meaning of Fiction, Detroit, 1960

Croll, Morris W. "The Baroque Style in Prose," Studies . . . in Honor of Frederick Klaeber, 1929

Davie, Donald. The Language of Science and the Language of Literature, 1700-1740, London, 1963

Davies, Hugh Sykes. "Trollope and his Style," Review of English Literature, I (1960), 73-85

Devoto, Giacomo. Linguistcs and Literary Criticism, trans. M. F. Edgerton, Jr, New York, 1963

Dobree, Bonamy. Modern Prose Style, London, 1934

Dostal, Joseph. "A Study of Sentences as to Usage and Length of Words," The English Journal, XV (1926), 145-147

Enkvist, Nils E. (In Spencer, John, see below)

Forster, E. M. Aspects of the Novel, London, 1927

Frye, Northrop. The Anatomy of Criticism, Princeton, 1957

Fulcher, Paul. Foundations of English Style, New York, 1927

Guiraud, Pierre (et al). Style and Litterature, The Hague, 1962

Kettle, Arnold. An Introduction to the English Novel. 2 vols. (Harper Torchbooks, TB1011, 1012).

Kubler, George. The Shape of Time, New Haven, 1962

Lacelles, Mary. Jane Austen and Her Art, London, 1939

Lesser, Simon D. Fiction and the Unconscious, New York, 1962

Liddell, Robert. Some Principles of Fiction, London, 1953

Lodge, David. Language of Fiction, New York, 1966

Lubbock, Percy. The Craft of Fiction, London, 1921

Lucas, F. L. Style, London, 1955

Martin, Harold C. Style in Prose Fiction, New York, 1959

Miles, Josephine. Eras and Modes in English Poetry, Berkeley & Los Angeles, 1963

Moirer, Henri. La psychologie des Styles, Geneva, 1954

Murray, J. Middleton. The Problem of Style, London, 1922

Ohmann, Richard. "Generative Grammar and the Concept of Literary Style," Word, XX (1964), 423-439

Ohmann, Richard. "Methods in the Study of Victorian Style," The Victorian Newsletter, 27 (1965), 1-4

Pascal, Roy. "Tense and Novel," The Modern Language Review, LVII (1962), 1-11

Pascal, Roy. "The Present Tense in 'The Pilgrim's Progress," The Modern Language Review, LX (1965), 13-16

Ross, John F. Swift and Defoe, Berkeley and Los Angeles, 1941

Sayce, R. A. "The Definition of the Term 'Style,'" Proceedings of the 3rd Congress of the International Comparative Literature Association, The Hague, 1962

Saintsbury, George. The English Novel, London, 1913

Saintsbury, George. A History of English Prose Rhythm, London, 1912

Scholes, Robert and Robert Kellogg. The Nature of Narrative, New York, 1966

Spencer, John, ed. Linguistics and Style, London, 1964 (includes Nils Erik Enkvist's Language and Style)

Stang, Richard. The Theory of the Novel in England, 1850-1870, New York, 1959

Sutherland, James. On English Prose, London, 1957

Thomsom, J.A.K. Classical Influences on English Prose, Longon, 1956

Ullman, Stephen. Semantics, Oxford, 1962

Ullman, Stephen. Language and Style, Oxford, 1964

Vincent, E. R. Literature and Science, Oxford, 1955

Williamson, George. The Senecan Amble (1951)

APPENDIX A

Listing of Novels Treated in Project and Parts Analyzed

The form of listing is as follows:

Number assigned for purposes of the project
Title of Novel with date of original publication in book form -- sometimes date of composition added
Author of Novel
Edition of novel utilized
Passages analyzed for grammar and vocabulary study

1
Emma (1816)
Austen, Jane
Houghton-Mifflin, Riverside edition, Lionel Trilling editor (B 17)
1. Random samples: 50 groups of five consecutive sentences selected by a table of random numbers applied to pages and then sentences on pages throughout the novel
2. Pages 190-196: Beginning of Chapter 11, Part II through all p. 196.

2
Northanger Abbey (1818)
Austen, Jane
Holt, Rinehart and Winston, with The Castle of Otranto and selections from The Mysteries of Udolpho, edited by Andrew Wright
1. Random samples: same system as used for Emma
2. Pages 411-421: beginning linell, p. 411, through line 20, p. 421; includes a portion of Chapter 15, Part I and a portion of Chapter 1, Part II.

3
Pride and Prejudice (1813)
Austen, Jane
Houghton-Mifflin, Riverside edition, Mark Schorer editor (B1)
Pages 126-136, Chapters 7,8,9, Volume II.

4
Persuasion (1818)
Austen, Jane
Norton, The Norton Library (N163), David Daiches editor
1. Pages 102-128 (Chapter 12, Volume I and Chapter I, Volume II) plus random samples of groups of five consecutive sentences each.
2. Pages 145-151, Chapter 4, Volume II.

5
Jacob's Room (1922)
Woolf, Virginia
Harcourt Brace and World, Harvest Books (HB37)
Pages 74-99, Chapters 6 through 8; also ten random samples as in *Emma*

6
Jane Eyre (1847)
Bronte, Charlotte
Houghton Mifflin, Riverside Editions (B 35), Mark Schorer editor
1. Pages 209-228, Chapter 21; 10 random samples as in *Emma*
2. Pages 299, line 5, through 305, complete

7
Middlemarch (1871-72)
Eliot, George
Houghton Mifflin, Riverside Editions (B 6), Gordon S. Haight editor
1. Pages 301-322, Chapters 41 through 44; also 10 random samples as in *Emma*
2. Pages 315-320, Chapter 43.

8
Joseph Andrews (1742)
Fielding, Henry
Houghton Mifflin, Riverside Editions (B 62), Martin C. Battestin editor
1. Pages 88-104, Chapters 4 and 5 of Book II
2. Pages 137-144, Chapters 14 and 15 of Book II.

9
Sons and Lovers (1913)
Lawrence, D. H.
Compass Books, The Viking Press
Pages 196, line 7, through page 208, line 2 all in Chapter 8.

10
Evelina (1778)
Burney, Fanny
Everyman Library (352)
Pages 179 through 191, Letter XLVI through Letter XLVII

11
Mansfield Park (1814)
Austen, Jane
Washington Square Press, Introduction Henry H. Adams
Pages 189-200, Chapters 24 and 25

12
Villette (1853)
Bronte, Charlotte
Dell, Laurel Edition
Page 262, line 28, through page 274, line 24 (part of Chapter 21 and part of 22)

13
Sense and Sensibility (1811)
Austen, Jane
Modern Library College Edition, introduction David Daiches
Pages 448-461, Chapters 26 and 27.

14
Wuthering Heights (1847)
Bronte, Emily
Houghton Mifflin Riverside Edition (B 2), V. S. Pritchett, editor
Pages 132-144, Chapters 15 and 16

15
The Mill on the Floss (1860)
Eliot, George
Houghton Mifflin Riverside Edition (B54), Gordon S. Haight, editor
Pages 224-236, Chapter 8, Book III and Chapter 9, Book III.

16
Adam Bede (1859)
Eliot, George
Holt, Rinehart and Winston, Gordon S. Haight editor
Pages 264-277, Chapters 23 and 24

17
Tom Jones (1749)
Fielding, Henry
Modern Library College Edition, introduction by George Sherburn
Pages 441-453, Chapter 7 Book IX through Chapter 2 Book X.

18
Great Expectations (1861)
Dickens, Charles
Doubleday Dolphin Book
Pages 250-267, Beginning of Chapter 29 through line 4, page 267, in Chapter 30.

19
Pamela (1740)
Richardson, Sammuel
Norton Library (N 166)
Pages 258, line 26, through page 270, line 36.

20
Nostromo (1904)
Conrad, Joseph
Modern Library, Introduction by Robert Penn Warren
No grammatical analysis

21
Oliver Twist (1838)
Dickens, Charles
Holt, Rinehart and Winston, introduction by J. Hillis Miller
Pages 201-213, Chapters 28 and 29

22
Daniel Deronda (1876)
Eliot, George
Harper Torchbook, The Academy Library (TB 1039), Introduction by F. R. Leavis
Pages 301-313, line 19, first part of Chapter 35, beginning of Book V.

23
David Copperfield (1850)
Dickens, Charles
Houghton Mifflin Riverside Edition (B 24), George H. Ford, editor
Pages 331-343, Chapters 29 and 30

24
Vanity Fair (1848)
Thacheray, William Makepeace
Houghton Mifflin Riverside Edition, edited by Geoffrey and Kathleen Tillotson
Pages 307, line 25, through Page 320, line 4, in Chapter 32 into Chapter 34

25
Agnes Grey (1847)
Bronte, Anne
Everyman Library (685), introduction by Margaret Lane
Pages 431, beginning of Chapter 6, through page 439, line 14, in Chapter 8.

26
The Tenant of Wildfell Hall (1848)
Bronte, Anne
Zodiac Press, London, 1954
Pages 208-217, Chapter 30.

27
Lady Susan (1871-c.1805)
Austen, Jane
The Works of Jane Austen, edited by R. W. Chapman, vol VI, Minor Works (Oxford)
Pages 272-279, Letters 18-21

28
Love and Friendship (Volume the Second) (1922 - before 1800)
Austen, Jane
The Works of Jane Austen, edited by R. W. Chapman, vol VI, Minor Works (Oxford)
Pages 82-88, Letters 7-9.

29
Romola (1863)
Eliot, George
Harpers Library Edition
Page 201-210, Chapters 23 and 24 not including page 206, line 3 through page 209, line 5 (Savanarola's sermon).

30
Old Mortality (1816)
Scott, Sir Walter
Everyman Library (137)
Pages 210-218, Chapter 21.

31
Far From the Madding Crowd (1874)
Hardy, Thomas
Houghton Mifflin, Riverside Edition (B 18) edited by Richard L. Purdy
Pages 165-171, Chapter 29, and pages 176, beginning Chapter 31, through 177, line 12.

32
Silas Marner (1861)
Eliot, George
Holt, Rinehart and Winston, introduction by Jerome Thale
Pages 106, line 24 through page 122, line 30, parts of Chapters 10 and 11, Part I.

33
Scenes from Clerical Life (1858)
Eliot, George
William Blackwood and Sons, Cabinet Edition, 24 volumes, Vol I.
Pages 66-72, in Chapter 5 "Amos Barton"; Pages 203-209, line 9, in Chapter 5, Mr. Gilfil's Love Story."

34
Felix Holt (1866)
Eliot, George
Everyman Library No. 353
Pages 206-213, Chapter 22

35
The Professor (1857)
Bronte, Charlotte
Everyman Library No. 322
Pages 114-120, Chapter 16

36
Shirley (1849)
Bronte, Charlotte
Everyman Library No. 288
Pages 251-261, Chapter 18

37
The Power and the Glory (1940)
Greene, Graham
Compass Books C 40, Viking Press
Pages 139, beginning Chapter 2 through page 143, line 22, and Pages 155, line 17 through page 158, line 19

38
Moll Flanders (1772)
Defoe, Daniel
Houghton Mifflin Riverside Edition
Pages 148, paragraph 2, through 160, paragraph 3.

39
The Green Dwarf (1933) (1833)
Bronte, Charlotte
Legends of Angria by Fannie E. Ratchford, New Haven, 1933
Pages 49, beginning Chapter 5, through page 56, line 33

40
Caroline Vernon (1933) (1839)
Bronte, Charlotte
Legends of Angria (see 39)
Pages 263, line 15, through 269, line 2

41
The Spell (1931) (1834)
The Spell: An Extravaganza, edited with an introduction by George Edwin MacLean, Oxford University Press, London, 1931
Pages 66, beginning Chapter 5, through 72, line 22

42
Bleak House (1853)
Dickens, Charles
Houghton Mifflin Riverside Edition (B 4) edited by Morton Dauwen Zabel
Pages 333, line 29, through 339, line 47, in Chapter 31 into Chapter 32

43
The Pickwick Papers (1837)
Dickens, Charles
Signet Classic, CQ200
No grammatical analysis

44
Our Mutual Friend (1865)
Dickens, Charles
Signet Classics CQ 244
Page 436, paragraph 7, through page 444, paragraph 1, in Chapter 32

45
Jude the Obscure (1894)
Hardy, Thomas
Dell, Laurel Edition
Pages 199, Beginning Chapter 10, through 206, line 6, in Chapter 1, Part IV

46
The Way We Live Now (1875)
Trollope, Anthony
World Classics No. 484
Pages 466, line 26 through 478 end of Chapter 50 (end Volume I)

47
The Portrait of a Lady (1881)
James, Henry
No grammatical analysis

48
What Maisie Knew (1897)
James, Henry
No grammatical analysis

49
The Ambassadors (1903)
James, Henry
No grammatical analysis

50
A Portrait of the Artist as a Young Man (1916)
Joyce, James
Compass Books C 9, Viking Press
Pages 124, paragraph 2, through 126, paragraph 8 and 136, paragraph 1, through page 141 *to* paragraph 1.

61
"*Gondal's Queen*" (1955) (1841-44) (*The Death of A.G.A.*)
Bronte, Emily
Gondal's Queen, Fannie E. Ratchford, Austin, Texas, 1955, pp. 143-153
Complete text, lines 1-344

62
The Spanish Gypsy (1868)
Eliot, George
Thomas Nelson Sons, *Works of George Eliot*, New York, 1915
Pages 174, line 22 through 185, line 20, omitting lyrics "The world is great" and "O bird that used to press"

APPENDIX B

MACRO-MARKING PROCEDURES

1. Preparing the graph paper

There are four "macro-marking" elements: Time, Action, Location, and Character. For each element sheets of graph paper are prepared by numbering in the left-hand margin so that each horizontal row of the graph paper corresponds with a page in the novel to be marked. Numbering is begun on sheet one (each sheet contains 50 rows divided by thickened green lines into 10 sections of 5 rows each) so that for page 5 (and multiples of five, 10, 15, 20 etc.) the number is marked opposite the row immediately above a thick green line. This system makes it possible to indicate the number of each fifth page only, that is, one doesn't have to write each page number in, though there is always a row for every page. If, for example, a novel begins not on page 1 but on page 8, one begins by writing an 8 in the left-hand margin of the first sheet on the third row below a thick green line. A thick pencil line is drawn through the row above the beginning row, leaving at least the first 20 horizontal lines on the first sheet blank, so that there is space to write element information in the columns above the first marked row--see instructions below. At the top of each sheet of paper is put the name and number of the novel at the center, and the page number of the sheet plus the name of the element at the far left. Thus

2/Location   23 David Copperfield

There is a complete set of sheets for each element. Thus, if it takes 8 sheets to provide enough rows to match the number of pages in the novel, one ends up with four sets (one for each element) of 8 sheets each.

Opposite the appropriate row (= page number) in the left-hand margin is indicated where each chapter, volume, or other division (sometimes a row of asterisks) occurs. Simple abbreviations, such as C for chapter followed by its number, or vol. for volume or Bk for book, are satisfactory as long as they are clear in meaning and indicate the page on which they occur (only beginnings are marked).

Markings always consist (except where exceptions are noted) of an X in the appropriate row and column of a sheet.

2. Element categories.

A. Location

A thick pencil line is marked a minimum of 20 rows from the top of the graph paper all the way across the sheet. The columns are numbered consecutively from 1 through 35. In column 1 is written INDEFINITE and in column 2 NEGATIVE vertically between the thick pencil line and the top of the sheet. Each column after 2 is used to identify each separate location that occurs in the novel in the order in which the locations occur. Each location is identified by writing a descriptive name in the appropriate column vertically between the thick line and the top of the page. The appropriate proper name is used where possible (London, not the city). If a descriptive term is necessary, it is made as precise as possible and appropriate. The grounds and immediate vicinity of a house may count as the house itself in some novels, but in others it may be important to distinguish between a house and its garden or grounds. Likewise in some novels all rooms in a house may be lumped under one location name, but in other novels separate rooms in the same house are distinguished. There may be more than 35 locations. If so, a new set of sheets is started numbered 36-70 across the top (of course, with the appropriate page numbers along the left margin). These sheets are numbered consecutively as "a" : 4a, 5a, 6a, etc. and placed immediately after the original sheet with locations 1-35.

B. Character

The procedure for marking Location is used for Character. The character name, written vertically in the appropriate column, should follow the ordinary mode of address, e.g. Sir Walter Elliot. If the character is not identified by name, he is identified by means of a descriptive phrase, e.g. "man who pours beer on Parson Adams." In the case of very minor charcters, each is assigned a number on a separate sheet of page, starting with 201 for the first such minor character, 202 for the second, etc., and, following the number, name (or identifying phrase), a list of the page numbers on which the minor character appears. A character is not given a column or placed on the special "minor character" sheet if he is only talked about and does not "appear." Abstractions and the "omniscient" author's voice are not marked (though of course in a first-person narrative the "author" is a character and must be given a column). A 1 instead of an X is marked in the appropriate column if a character is present in a scene but not participating, that is, not speaking or not having his actions narrated (this situation occurs usually only in large party scenes).

C. Time

Fifteen columns are needed for marking this element. The columns are labelled in the following manner:

col 1: minutes
2: hours
3: days
4: weeks
5: months
6: years
7: seconds

col 8: indeterminate
9: negative
10, 11, 12: "meanwhile"
13: flashback
14: flashforward
15: multiple flashbacks

Columns refer to the amount of time "covered" by the action on the page to which each particular row refers. Several time periods may occur on a single page, e.g., a conversation may be followed by a rapid narrative summary of several weeks and this in turn may be followed by narration of a few hours' activities: in this case an X would be placed in columns 1, 2, and 4. Conversation is ordinarily marked as "minutes" under column 1 unless there is special reason for marking it otherwise. Column 9, "negative" is used only when there is no "action" of the novel-plot itself, as when as author philosophizes (again, in a first-person narrative, however, such philosophizing would probably not be marked as negative). Column 8, indeterminate, is self-explanatory, but is used sparingly. The "meanwhile" columns (10-12) are most likely to be used in a novel with multiple plots, for they indicate situations in which an action occurs simultaneously with another action (usually previously narrated, but in some modern novels, particularly, simultaneity may be more complex). The marking of such simultaneity in the "meanwhile" columns is handled as follows: when one reaches a point where the time is simultaneous with a previously narrated time, one goes back to the original occurrence of this "simultaneous time" and marks for all pages involved a zero in column 11 (and a 1 in column 12); then one marks all pages for the "second" occurrence of this time, the pages dealing with action that occurs simultaneously with the "original" time, with an "A" in column 10 and a zero in column 11 and a 1 in column 12; if later there is a third simultaneous action, one marks it like the second but with a B in column 10 and a zero in column 11 and a 1 in column 12. The next "set" of simultaneous actions is marked in the same way except that an 02 is marked in column 11 and 12; the third set requires an 03 in columns 11 and 12, etc. An X is marked in column 13, flashback, for any reference of over two sentences to the past; the same applies to column 14, flashforward. Column 15 is marked with an X whenever there is a combination of flashbacks within flashbacks, flashforwards with flashforwards, or flashbacks occur with flashforwards. In other words, column 15 indicates there is more than a single "distortion" of the time sequence. When one marks column 15 one also marks either column 13 or column 14 (or both) depending on which is appropriate, e.g. a flashback within a flashback requires an X in columns 13 and 15 but not in column 14.

D. Action

Action is marked in the same way as time, but there are only seven regular categories of action:

col 1: narration
col 2: description
col 3: exposition
col 4: "interiorized"
col 5: dialogue
col 6: quoted letter, poem, etc.
col 7: narrative within narrative
col 8: "other"

Column 1, narration, is the most commonly marked category, and in many novels is marked for almost every page. If a page is nearly all dialogue, however, except for such bits of narration as "She thought-fully said . . ." --is not marked. The same is true in reverse--if only half a sentence of dialogue appears on a page, it is not marked the dialogue column, 5. But, as with "Time," markings in more than one column for a page are not uncommon, and in some cases all the "Action" columns will be marked for a single page. "Exposition," column 3, means the comments by an omniscient author (this column, for example, is not infrequently marked in George Eliot's novels). Description, column 2, is usually distinguishable from narration, but not infrequently narration and description are intermingled, in which case both columns 1 and 2 are marked. Column 4, "interiorized" refers to the representation of psychic actions, the thoughts of a character, his soliloquies, and so forth. This column is marked when there is a "stream-of-consciousness" passage, for example, and also when there is a formal soliloquy presented in quotation marks. Column 6 refers to quotations and also to the reproduction of letters and the like, when the novel is not epistolary in form. Notice that in some cases one marks both columns 6 and 7, when a long letter tells a story, for example. Column 7 is used for all narratives within narratives, or narratives within narratives within narratives, etc. Again, a mark in column 7 may be accompanied by marks in other columns. Column 8 is for situations which may arise not covered by columns 1-7.

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO.

CLEARINGHOUSE ACCESSION NUMBER

RESUME DATE 30–10–66

P.A.

T.A.

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒

ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

TITLE

A Computer Analysis of Fictional Prose Style: Final Report

PERSONAL AUTHOR(S)

Kroeber, Karl

INSTITUTION (SOURCE)

Department of English, University of Wisconsin, Madison, Wis. 53706

SOURCE CODE

REPORT/SERIES NO.

OTHER SOURCE

SOURCE CODE

OTHER REPORT NO.

OTHER SOURCE

SOURCE CODE

OTHER REPORT NO.

PUB'L. DATE 30 Oct. –66

CONTRACT/GRANT NUMBER OE -6-10-015

PAGINATION, ETC.

120 pages; Project No. 3067

RETRIEVAL TERMS

fiction; prose style; syntax; vocabulary; computer-analysis of language; grammatical constructions; parts-of-speech; sentence length; sentence structure; nineteenth-century novelists; fictional language; novel; style; word frequencies; teaching of literature

IDENTIFIERS

Jane Austen; Brontës; George Eliot

ABSTRACT

This project tried to define in systematic and objective fashion fundamental characteristics of fictional prose style through analysis of syntax and vocabulary. In this process a large amount of information (never before collected) about the language used in novels was accumulated. Sample passages from all major novels of Jane Austen, the Brontë sisters, and George Eliot (all nineteenth-century novelists) as well as from novels by thirteen other authors from Daniel Defoe to Graham Greene were analyzed syntactically. Information about each sentence, clause, and word was coded, keypunched on IBM cards, and transferred to magnetic tape. Programs were devised for a CDC 3600 computer to run statistical tests on this data and to print out vocabulary lists with frequencies. Too much data (250,000 records) was collected to be fully analyzed quickly, but tentative results point to these conclusions: parts of speech and some grammatical constructions are, within narrow ranges of variation, consistent in all novelists and are not the best stylistic indicators; the data does permit some long-term trends in fictional prose to be identified through grammatical form; sentence-length and sentence-structure ratios between dialog and narrative prose appear to be potentially useful descriptors of novelistic style; larger samples analyzed automatically will be requisite for further advances in this research, as will more sophisticated statistical techniques of measurement; definition of structural elements of novels as total units should precede and determine detailed syntactic and vocabulary analyses. The hypotheses and techniques of this work, on the basis of one teaching experiment, appear useful as a basis for innovations in the teaching of literature.

OE 6000 (Rev. 9-66)

INSTRUCTIONS FOR COMPLETING ERIC REPORT RESUME

The resume is used to identify summary data and information about each document acquired, processed, and stored within the ERIC system. In addition to serving as a permanent record of the document in the collection, the resume is also a means of dissemination. All fields of the form must be completed in the allotted spaces, but inapplicable fields should be left blank. The following instructions are keyed to the line numbers appearing in the left margin of the form:

TOP LINE. ERIC Accession No. Leave blank. A permanent ED number will be assigned to each resume and its corresponding document as they are processed into the ERIC system.

LINE 001. Clearinghouse Accession No. For use only by ERIC Clearinghouses. Enter the alpha code and 6-digit document number.

Resume Date. In numeric form, enter month, day, and year that resume is completed. *(Example: 07 14 66)*

P.A. Leave blank.

T.A. Leave blank.

Copyright. Check appropriate block to denote presence of copyrighted material within the document.

ERIC Reproduction Release. Check appropriate block to indicate that ERIC has permission to reproduce the document and its resume form.

LINES 100-103. Title. Enter the complete document title, including subtitles if they add significant information. Where applicable, also enter volume number or part number, and the type of document *(Final Report, Interim Report, Thesis, etc.)*.

LINE 200. Personal Author(s). Enter personal author(s), last name first. *(Example: Doe, John J.)* If two authors are given, enter both. *(Example: Doe, John J. Smith, Ted)*. If there are three or more authors, list only one followed by "and others."

LINE 300. Institution *(Source)*. Enter the name of the organization which originated the report. Include the address *(city and State)* and the subordinate unit of the organization. *(Example: Harvard Univ., Cambridge, Mass., School of Education.)*

Source Code. Leave blank.

LINE 310. Report/Series No. Enter any unique number assigned to the document by the institutional source. *(Example: SC-1234)*

LINE 320. Other Source. Use only when a second source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 330. Other Report No. Enter document number assigned by the second source.

LINE 340. Other Source. Use only when a third source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 350. Other Report No. Enter document number assigned by the third source.

LINE 400. Publication Date. Enter the day, month, and year of the document. *(Example: 12 Jun 66)*

Contract/Grant Number. Applicable only for documents generated from research sponsored by the U.S. Office of Education. Enter appropriate contract or grant number and its prefix. *(Example: OEC-1-6-061234-0033)*

LINES 500-501. Pagination, etc. Enter the total number of pages of the document, including illustrations and appendixes. *(Example: 115p.)* USE THIS SPACE FOR ADDITIONAL INFORMATION PERTINENT TO THE DOCUMENT, such as publisher, journal citation, and other contract numbers.

LINES 600-606. Retrieval Terms. Enter the important subject terms *(descriptors)* which, taken as a group, adequately describe the contents of the document.

LINE 607. Identifiers. Enter any additional important terms, more specific than descriptors, such as trade names, equipment model names and numbers, organization and project names, discussed in the document.

LINES 800-822. Abstract. Enter an informative abstract of the document. Its style and content must be suitable for public announcement and dissemination.

U.S. GOVERNMENT PRINTING OFFICE : 1966 O—221-551